

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**THIS PAGE BLANK (USPTO)**



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 108626**

**TO: Nita M Minnifield**  
**Location: cm1/8a07/8e12**  
**Art Unit: 1645**  
**Tuesday, November 25, 2003**  
**Case Serial Number: 09/298523**

**From: Paul Schulwitz**  
**Location: Biotech-Chem Library**  
**CM1-6B06**  
**Phone: 305-1954**

**paul.schulwitz@uspto.gov**

### **Search Notes**

Examiner Minnifield,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz  
Technical Information Specialist  
STIC Biotech/Chem Library  
(703)305-1954

**THIS PAGE BLANK (USPTO)**



108626

From: Chan, Christina  
Sent: Wednesday, November 19, 2003 11:51 AM  
To: Minnifield, Nita; STIC-Biotech/ChemLib  
Subject: RE: rush sequence search request

RECEIVED

NOV 19 2003

STIC-BIOTECH/ChemLib  
(STIC)

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

-----Original Message-----

From: Minnifield, Nita  
Sent: Wednesday, November 19, 2003 11:30 AM  
To: Chan, Christina  
Subject: rush sequence search request

Christina, please approve, 2 month amdt.

STIC  
09/298523

Please do a commercial and interference sequence search on SEQ ID  
NO: 13 and aa 263-442 of SEQ ID NO: 13.

Please provide a paper copy of the results.

Thanks,

Nita M. Minnifield  
Art Unit 1645  
Office CM1-8A07  
Mailbox CM1-8E12  
703-305-3394

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/21  
Date Completed: 11/25  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 21.4457 Seconds  
(without alignments)  
1008.166 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFAKSERKHYHSIRKFSVG.....NRLNQOPPKTEKPAQSTP 511

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/aa/6C.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/aa/6D.COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194.5	86.5	446	US-09-286-981B-9	Sequence 9, Appli
2	2177.5	85.8	446	US-09-286-981B-6	Sequence 6, Appli
3	1728	68.1	564	US-09-308-022-6	Sequence 6, Appli
4	1512.5	59.6	406	US-09-286-981B-18	Sequence 18, Appli
5	1506	59.4	631	US-08-847-065-25	Sequence 25, Appli
6	1455.5	57.4	426	US-09-286-981B-12	Sequence 12, Appli
7	1447.5	57.0	425	US-09-286-981B-14	Sequence 14, Appli
8	1446.5	57.0	424	US-09-286-981B-13	Sequence 13, Appli
9	1440.5	56.8	428	US-09-286-981B-7	Sequence 7, Appli
10	1432	56.4	425	US-09-286-981B-11	Sequence 11, Appli
11	1412.5	55.7	412	US-09-286-981B-17	Sequence 17, Appli
12	1410.5	55.6	419	US-09-286-981B-15	Sequence 15, Appli
13	1405.5	55.4	414	US-09-286-981B-16	Sequence 16, Appli
14	1401.5	55.2	414	US-09-286-981B-10	Sequence 10, Appli
15	1395	55.0	414	US-09-286-981B-3	Sequence 3, Appli
16	1373	54.1	453	US-08-961-083-38	Sequence 38, Appli
17	1373	54.1	453	US-09-538-784-38	Sequence 38, Appli
18	1291	50.9	413	US-09-286-981B-5	Sequence 5, Appli
19	1207	47.6	864	US-08-714-741-40	Sequence 40, Appli
20	1157	45.6	588	US-08-714-741-42	Sequence 42, Appli
21	904	35.6	605	US-08-714-741-46	Sequence 46, Appli
22	856.5	33.8	8991	US-08-714-741-32	Sequence 32, Appli
23	764	30.1	1231	US-08-714-741-41	Sequence 41, Appli
24	752	29.6	623	US-08-714-741-47	Sequence 47, Appli
25	714.5	28.2	251	US-08-286-981B-4	Sequence 4, Appli
26	636	25.1	142	US-08-847-065-21	Sequence 21, Appli
27	560	22.1	219	US-09-286-981B-8	Sequence 8, Appli

28	542	21.4	114	4	US-09-286-981B-19	Sequence 19, Appli
29	525.5	20.7	110	3	US-08-961-083-102	Sequence 102, App
30	525.5	20.7	110	4	US-09-536-784-102	Sequence 102, App
31	499.5	19.7	695	1	US-08-127-499A-23	Sequence 23, Appli
32	499.5	19.7	695	1	US-08-482-847-23	Sequence 23, Appli
33	491	19.4	619	1	US-08-465-746-2	Sequence 2, Appli
34	491	19.4	619	1	US-08-214-164-2	Sequence 2, Appli
35	491	19.4	619	2	US-08-467-852A-3	Sequence 3, Appli
36	491	19.4	619	2	US-08-246-636-2	Sequence 2, Appli
37	491	19.4	619	2	US-08-247-491A-3	Sequence 3, Appli
38	491	19.4	619	2	US-08-319-795-2	Sequence 2, Appli
39	491	19.4	619	2	US-08-468-985-2	Sequence 2, Appli
40	491	19.4	619	3	US-08-312-949-2	Sequence 2, Appli
41	491	19.4	648	1	US-08-072-070-2	Sequence 2, Appli
42	491	19.4	648	1	US-08-469-434-2	Sequence 2, Appli
43	491	19.4	648	1	US-08-214-222-2	Sequence 2, Appli
44	491	19.4	648	2	US-08-467-852A-2	Sequence 2, Appli
45	491	19.4	648	2	US-08-468-718-2	Sequence 2, Appli

#### ALIGNMENTS

```
RESULT 1
US-09-286-981B-9
Sequence 9, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizenmann, Theresa M.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286, 981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085, 743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-9
Query Match 86.5%; Score 2194.5; DB 4; Length 446;
Best Local Similarity 99.8%; Pred. No. 7.3e-146;
Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 37 TENEGATQVPTSSNRANESQAEOGEPKUDSRDKARKVEEYVKIIGESTAKTKKR 96
1 TENEGATQVPTSSNRANESQAEOGEPKUDSRDKARKVEEYVKIIGESTAKTKKR 60
DQ 97 HTTVALVNLNNIKNEYLNKIVESTSESQLOIMESRERKVNBAVAFKFDSSSSSSD 156
HTTVALVNLNNIKNEYLNKIVESTSESQLOIMESRERKVNBAVAFKFDSSSSSSD 120
DQ 61 HTTVALVNLNNIKNEYLNKIVESTSESQLOIMESRERKVNBAVAFKFDSSSSSSD 120
HTTVALVNLNNIKNEYLNKIVESTSESQLOIMESRERKVNBAVAFKFDSSSSSSD 120
DQ 157 SSTPEASDTAKNKPTEPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITTYKTLEI 215
SSTPEASDTAKNKPTEPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITTYKTLEI 180
DQ 121 SSTPEASDTAKNKPTEPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITTYKTLEI 180
SSTPEASDTAKNKPTEPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITTYKTLEI 180
DQ 216 AESDVEYKAEELVYKKNAPRDEOKIKOAEVSKOAEARLKKIKTDREAEAEAK 275
AESDVEYKAEELVYKKNAPRDEOKIKOAEVSKOAEARLKKIKTDREAEAEAK 240
DQ 181 AESDVEYKAEELVYKKNAPRDEOKIKOAEVSKOAEARLKKIKTDREAEAEAK 240
AESDVEYKAEELVYKKNAPRDEOKIKOAEVSKOAEARLKKIKTDREAEAEAK 240
DQ 276 RRADAEQKPKGRKRGVGEIATPDKKENDAKSSSVGEETLPSPSLKPEKVAEAE 335
RRADAEQKPKGRKRGVGEIATPDKKENDAKSSSVGEETLPSPSLKPEKVAEAE 300
DQ 241 RRADAEQKPKGRKRGVGEIATPDKKENDAKSSSVGEETLPSPSLKPEKVAEAE 300
RRADAEQKPKGRKRGVGEIATPDKKENDAKSSSVGEETLPSPSLKPEKVAEAE 300
DQ 336 KKYVEAKKKAEDQKEEDRRNYPITTYKTLEIAESDVEYKAEELVYKKEAEPRNEEK 395
KKYVEAKKKAEDQKEEDRRNYPITTYKTLEIAESDVEYKAEELVYKKEAEPRNEEK 360
DQ 301 KKYVEAKKKAEDQKEEDRRNYPITTYKTLEIAESDVEYKAEELVYKKEAEPRNEEK 360
KKYVEAKKKAEDQKEEDRRNYPITTYKTLEIAESDVEYKAEELVYKKEAEPRNEEK 360
```

```

1 APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
2 TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
3 TITLE OF INVENTION: PNEUMONIAE
4 NUMBER OF SEQUENCES: 12
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
7 STREET: 119 No. 6291654th Fourth Street, Suite 203
8 CITY: Minneapolis
9 STATE: Minnesota
10 COUNTRY: USA
11 ZIP: 55401
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentin Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/308,022
20 FILING DATE:
21
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: PCT/US97/20586
25 FILING DATE: 12-NOV-1997
26 APPLICATION NUMBER: 60/029,444
27 FILING DATE: 12-NOV-1996
28 APPLICATION NUMBER: 60/038,086
29 FILING DATE: 18-FEB-1997
30 APPLICATION NUMBER: 60/059,368
31 FILING DATE: 19-SEP-1997
32 APPLICATION NUMBER: 60/062,473
33 FILING DATE: 16-OCT-1997
34 ATTORNEY/AGENT INFORMATION:
35 NAME: MUETING, Ann M.
36 REGISTRATION NUMBER: 33,977
37 REFERENCE/DOCKET NUMBER: 110,00430101
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 612-305-1217
40 TELEFAX: 612-305-1228
41 INFORMATION FOR SEQ ID NO: 6:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 564 amino acids
44 TYPE: amino acid
45 STRANDEDNESS: single
46 TOPOLOGY: linear
47 MOLECULE TYPE: protein
48
49 US-09-308-022-6
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
9
```

Db 259 KGRKRGVPGELATPDCKENDAKSSDSVGEETLPSLSKGVAAEAKVEAEKAK 318  
Qy 347 DQKEDRRNYPNTYKTTLELEIAESDVYKKAELVYVEAEPRNEEVKQAKAIVESK 406  
Db 319 DQKEDRRNYPNTYKTTLELEIAESDVYKKAELVYVEAEPRNEEVKQAKAIVESK 378  
Qy 407 KAEATRLERIKITDRKKAEEAEAKAEDVYKKEKAEPOAPAPAKAEKAPAPAKENPA 466  
Db 379 KAEATRLERIKITDRKKAEEAEAKAEDVYKKEKAEPOAPAPAKAEKAPAPAKENPA 436  
Qy 467 EOPKAEKADQAEEDYARRSEEEYNRLTQOQPPTEKPAOPSTP 511  
Db 437 EOPKAEKADQAEEDYARRSEEEYNRLTQOQPPTEKPAOPSTP 481

## RESULT 4

US-09-286-981B-18  
; Sequence 18, Application US/09286981B  
; Patent No. 6503511  
; GENERAL INFORMATION:  
; APPLICANT: Mizemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/09/286,981B  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 406  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-18

Query Match 59.6%; Score 1512.5; DB 4; Length 406;  
Best Local Similarity 73.0%; Pred. No. 2.9e-98;  
Matches 327; Conservative 28; Mismatches 46; Indels 47; Gaps 8;  
Qy 37 TENEGATQVPTSNRANESQAEQEQPKLDSEBDKAEVEEYVKIAGESVAKSTYKR 96  
Db 1 TENEGTQAPTSSNRGNSQA-----EHMKAKAKQVDEVIEKML-----QDNRK 44  
Qy 97 HTTVALVNLNINIKNEYLNKIIVSTSESQIILMESRSKVDAVSKFEKSSSSSSD 156  
Db 45 HTQVNGLLTKLGAIKTEYLRGLSVSKSEKSTAE-LPSEIKETLTAAPKQFKK----- 94  
Qy 157 SSTPEASDTAKPKPTPEPGKVAEAKKVEEA-KKAKDQKEEDRRNYPNTYKTTLEI 215  
Db 95 -----DLTKPK-----KVAEAEKVAEAKKADQKEEDRRNYPNTYKTTLEI 140  
Qy 216 AESDVYKKAELVYKVAKEPRDEQIKQAEAVESKQAEATRLKIKTDRKEAEAEK 275  
Db 141 AESDVYKKAELVYKVAKEPRDEQIKQAEAVESKQAEATRLKIKTDRKEAEAEK 200  
Qy 276 RRADAKQG---KPKGRKRGVPGELATPDCKENDAKSSDSVGEETLPSPLKPKKV 331  
Db 201 RRVDAKEQDESSKRSKRVKRGDLGEQATPDCKENDAKSSDSVGEETLPSPLKPKKV 260  
Qy 332 AEAETKVEAEKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKKAELVYVEAEKPR 391  
Db 261 AEAETKVEAEKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKKAELVYVEAEKPR 320  
Qy 392 NEEKVQAKAEVSKKAATRLKIKTDRKKAEEAEAKAEDVYKKEKAEPOAPAPAP 451  
Db 321 NEEKVQAKAEVSKKAATRLKIKTDRKKAEEAEAKAEDVYKKEKAEPOAPAPAP 380  
Qy 452 KAEKPAAPKPRNPAEOPKAEKPADQQA 479  
Db 381 QPEK--PAPKPRNPAEOPKAEKPADQQA 406

## RESULT 5

US-08-847-065-25  
; Sequence 25, Application US/08847065  
; Patent No. 6245335  
; GENERAL INFORMATION:  
; APPLICANT: Masure, H. Robert  
; APPLICANT: Rosencow, Carsten I.  
; APPLICANT: Tuomanen, Elaine  
; APPLICANT: Mizemann, Theresa M.  
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: FLOOR  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/847,065  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 631 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: N-terminal  
US-08-847-065-25

Query Match 59.4%; Score 1506; DB 3; Length 631;  
Best Local Similarity 73.5%; Pred. No. 1.4e-97;  
Matches 316; Conservative 39; Mismatches 35; Indels 40; Gaps 6;  
Qy 94 KKRHTTVALVNLNINIKNEYLNK--IVESTSESQIILMESRSKVDAVSKFEKSSS 151  
Db 9 RRRHTQVNLNINILSNIKTYLRLVNLVLEKSKDELPE---SELKATLDAEFKFKDT-- 63  
Qy 152 SSSSDSYTPASDTAKPKPTPEPGKVAEAKKVEEA-KKAKDQKEEDRRNYPNTYKT 210  
Db 64 -----LKPGEKVAEAKKVEAEKKAADQKEEDRRNYPNTYKT 102  
Qy 211 LELEIAESDVYKKAELVYKVAKEPRDEQIKQAEAVESKQAEATRLKIKTDRKEA 270  
Db 103 LELEIAESDVYKKAELVYKVAKEPRDEQIKQAEAVESKQAEATRLKIKTDRKEA 162  
Qy 271 EBEAKRRADAK-----EQKPKGRKRGVPGELATPDCKENDAKSSDSVGEETLP 321  
Db 163 EBEAKRRADAKLEAVATSDQKPKGRKRGVPGELATPDCKENDAKSSDSVGEETLP 222  
Qy 322 SPSLKEKKVAEAEKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKKAEL 381  
Db 223 SSSLKSGKVAEAEKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKKAEL 282  
Qy 382 LVVEAEKPRNEKVKQAKAEVSKKAATRLKIKTDRKKAEEAEAKAEDVYKKEK 441



SEQ ID NO 14  
LENGTH: 424  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-14

Query Match 57.0%; Score 1446.5; DB 4; Length 424;  
Best Local Similarity 68.5%; Pred. No. 1.2e-93;  
Matches 317; Conservative 40; Mismatches 47; Indels 59; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEOGQPKLDSERDARKVEVEYKIVGESYAKSTKR 96  
DB 1 TEKEVTTQVATSSNRANESQA-----TEHAKAKQVDEYIKKTL-----OLDRRK 44  
QY 97 HTITVALVNLNNIKNEYLNKIVESTSESQLOILMMESRSKYDEAVSKPEKSSSSSSSD 156  
DB 45 HTQVGLTKLGIVITETYLHGLSVSKKSSAE-LPSETIAKIDAAFEQPKDT----- 96  
QY 157 SSTKEASDTAKPNKPTERGKVAEAKKKVEEA-KKADQKEEDRRNYPTIYKTLLEI 215  
DB 97 -----LPTERGKVAEAKKKVEEA-KKADQKEEDRRNYPTIYKTLLEI 142  
QY 216 AESDVEVKKAELELVKVAENPRDQKIKQAEAVESQAEATRLKIKITDREAEAEAK 275  
DB 143 AESDVEVKKAELELVKVAEKESRDEKIKQAEAVESQAEATRLKIKITDREAEAK 201  
QY 276 RRADAK-----EQQPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSBLK 326  
DB 202 RRADAKLEAVANVASEQKSRARAEVLGELATPDKKENDAKSSDSSVGEETLPSBLK 261  
QY 327 PEKVAEAEKVEAEAKKAEADQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVKEE 386  
DB 262 PEKVAEAEKVEAEAKKAEADQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVKEE 321  
QY 387 AKERNNEKIKQAEVSKKAEATRLKIKITDRKA-EEBAKKAABEDYKKEKPAEQ 445  
DB 322 AKERNNEKIKQAEVSKKAEATRLKIKITDRKA-EEBAKKAABEDYKKEKPAEQ 381  
QY 446 QAPAPKAEK-----PAPAPKPNPAOPKAEKPADQA 479  
DB 382 QAPAPKAEKPTPEBPENPAPAPKPNPAOPKAEKPADQA 424

RESULT 9  
US-09-286-981B-7

Sequence 7, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT FILING DATE: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-7

Query Match 56.8%; Score 1440.5; DB 4; Length 428;  
Best Local Similarity 69.7%; Pred. No. 3.3e-93;  
Matches 311; Conservative 35; Mismatches 53; Indels 47; Gaps 7;  
QY 60 GEQPKLDSERDARK--EVEEYVKIVGESYAKSTKRHTITVALVNLNNIKNEYLN 116  
DB 6 GNNSTVTSQDLSKVADEVESHOSILKDVNKLKKVOHTQVADPNKLSKIKTKYLY 65

QY 117 K-IVESTSESQLOILMMESRSKYDEAVSKPEKSSSSSSDSTKPEASDTAKENKPT 174  
DB 66 ELNVLEEKSEAELELSKTKETKEBELTAPEQPKDTLS-----TE 104  
QY 175 PEKVAEAKKKVEEA-KKAKQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVK 233  
DB 105 PEKVAEAKKKVEEA-KKAKQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVK 164  
QY 234 ANEPRDEKIKQAEAVESQAEATRLKIKITDREAE-----EAKRRADAK 282  
DB 165 ANEPRDEKIKQAEAVESQAEATRLKIKITDREAE-----EAKRRADAK 224  
QY 283 QGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSBLPEKVAEAEKVEBAK 342  
DB 225 E-FKKTKRGVLGEPAITPDKKENDAKSSDSSVGEETLPSBLPEKVAEAEKVEBAK 282  
QY 343 KKAEDQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVKAEKERNNEKVAQAE 402  
DB 283 KKAEDQKEEDRRNYPTIYKTLLEIAESDVEVKKAELELVKAEKERNNEKVAQAE 342  
QY 403 VESKAEATRLKIKITDRKAEAEAKKAAAEEDYKKEKPAEQPA-----PAPKAE 455  
DB 343 VESKAEATRLKIKITDRKAEAEAKKAAAEEDYKKEKPAEQPA-----PAPKAE 402  
QY 456 PAPAPKPNPAOPKAEKPADQAE 481  
DB 403 PAPAPKPNPAOPKAEKPADQAE 428

RESULT 10  
US-09-286-981B-11

Sequence 11, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT FILING DATE: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 11  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-11

Query Match 56.4%; Score 1432; DB 4; Length 425;  
Best Local Similarity 66.6%; Pred. No. 1.3e-92;  
Matches 311; Conservative 44; Mismatches 48; Indels 64; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEOGQPKLDSERDARKVEVEYKIVGESYAKSTKR 96  
DB 1 TEKEVTTQVATSSNRANESQA-----KAAEQFDEYIKTMI-----OLDRRK 44  
QY 97 HTITVALVNLNNIKNEYLNK-IVESTSESQLOILMMESRSKYDEAVSKPEKSSSSSS 154  
DB 45 HTQVGLTKLGIVITETYLHGLSVSKKAELE-LPSETIAKIDAAFEQPKDTNRTTK 101  
QY 155 SDSSTKPEASDTAKPNKPTERGKVAEAKKKVEEA-KKADQKEEDRRNYPTIYKTLLEI 213  
DB 102 T-----VAEAEKVEEA-KKAKQKEEDRRNYPTIYKTLLEI 138  
QY 214 EIASDVEVKKAELELVKVAENPRDQKIKQAEAVESQAEATRLKIKITDREAE 273  
DB 139 EIASDVEVKKAELELVKVAEKESRDEKIKQAEAVESQAEATRLKIKITDREAE 198  
QY 274 AKRRADAK-----EQQPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 321

```
Db 199 AKRBAEAKLKEAVEKNVATSEODKRGKRKRGVPGEQATPPKENDAKSSDSSVGEALP 258
Qy 322 SPSLPEKKVAEAEKKVAEAEKKAEADQKEEDRRNPTTYKTLELEIESVYKKALE 381
Db 259 SPSLPEKKVAEAEKKVAEAEKKAEADQKEEDRRNPTTYKTLELEIESVYKKALE 318
Qy 382 LVKEAKEPRBEKKVAEAEKKAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 441
Db 319 LVKEAKEPRBEKKVAEAEKKAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 378
Qy 442 AEQOPAPAPAEK-----PAPAKPENPAEQPKAEKPADQAE 481
Db 379 AEQOPAPAPAEKTEPEENPAPAPAEKPADQAEKTDQAE 425

RESULT 11
US-09-286-981B-17
; Sequence 17, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 17
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-17

Query Match 55.7%; Score 1412.5; DB 4; Length 412;
Best Local Similarity 70.8%; Pred. No. 2.9e-91;
Matches 303; Conservative 41; Mismatches 51; Indels 33; Gaps 6;

Qy 63 PKLDSERDKARK--EVEEYVKIIGESYAKSTKRTITVALVNEINNIKNETLAK-- 117
Db 9 PTYSSGGDISKYADVKSHLEKLSIQNLDRSKIKITVNLINKQIDIKRTLYLNL 68
Qy 118 IVESITSESOQLIMMSSRSKVDEAVSKTEKSSSSSSSSSTKPEASDTAKPNKTEPGE 177
Db 69 VLEDKSKAELP---SKIKAEILDAAFEQPKDT-----LPTBPGK 104
Qy 178 KVAEAKKKVVEA--KRAKQKEEDRRNPTTYKTLELEIESDVYKKALELVKANE 236
Db 105 KVAEAKKKVVEA--KRAKQKEEDRRNPTTYKTLELEIESDVYKKALELVKANE 164
Qy 237 PRDEOKIKQAEAVESKQAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 293
Db 165 SRNEGTINQAKAVESQAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 224
Qy 294 VPEGLATPDKKENDAKSSDSSVGEETLPSLPEKKVAEAEKKVAEAEKKAEEDVKKEP 353
Db 225 DFGEPATPDKKENDAKSSDSSVGEETLPSLPEKKVAEAEKKVAEAEKKAEEDVKKEP 284
Qy 354 RNYPTNYKTLELEIESDVYKKALELVKAEAKPENBEKKVAEAEKKAEATRL 413
Db 285 RNYPTTYKTLELEIESDVYKKALELVKAEAGSNEBEKKVAEAEKKAEATRL 344
Qy 414 EKITDRKAAEEBKKAAEEDVKKEPAPQAPAPAEKPAEPAPAEQPKAEK 473
Db 345 EKITDRKAAEEBKKAAEEDVKKEPAPQAPAPAEKPAEPAPAEQPKAEK 404
Qy 474 PADQAE 481
Db 405 PADQAE 412
```

```
RESULT 12
US-09-286-981B-15
; Sequence 15, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-15

Query Match 55.6%; Score 1410.5; DB 4; Length 419;
Best Local Similarity 66.5%; Pred. No. 4e-91;
Matches 308; Conservative 43; Mismatches 49; Indels 63; Gaps 9;

Qy 37 TENEGATQVPTSSNRANESQAEQEPKLDSEDKARKEVEEYVKIIGESYAKSTKR 96
Db 1 TENERTQVPTSSNRGR-----PERKKAEPDEYINQI-----QLDKRK 41
Qy 97 HTITVALVNEINNIKNETLAKIIVESTESQQLIMMSSRSKVDEAVSKTEKSSSSSSSD 156
Db 42 HTQWLAFTIQLSRKITYLNGLKK--SEALP---SKIKAEILDAAFEQPKDT----- 90
Qy 157 SSTPEASDTAKPNKTEPEEKVAEAKKVEA-----KKAQKQEDRRNPTTY 208
Db 91 -----LPTPEKKVAEAEKKVAEAEKKVAEAEKKAEADQKEEDRRNPTTY 136
Qy 209 KTLLELEIESDVYKKALELVKANEPRDEOKIKQAEAVESKQAEATRLKIKITDR 268
Db 137 KTLLELEIESDVYKKALELVKANEPRDEOKIKQAEAVESKQAEATRLKIKITDR 196
Qy 269 EA--BEAKRRADAEQ--KPKGRKRGVPEGLATPDKKENDAKSSDSSVGEETLPS 324
Db 197 KAEBEAKRRADAEQDESKRKRSGALGEQATPDKKENDAKSSDSSVGEETLPS 256
Qy 325 LKPEKVAEAEKKVAEAEKKAEADQKEEDRRNPTTYKTLELEIESDVYKKALELVK 384
Db 257 LKPKVAEAEKKVAEAEKKAEADQKEEDRRNPTTYKTLELEIESDVYKKALELVK 316
Qy 385 EEAKEPRBEKKVAEAEKKAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 444
Db 317 EEAKEPRBEKKVAEAEKKAEATRLKIKITDRKAAEEBKKAAEEDVKKEP 376
Qy 445 POPAPAEKPA-----PAPAKPENPAEQPKAEKPADQAE 480
Db 377 POPAPAEKPAEPPEENPAPAPAEKPADQAE 419

RESULT 13
US-09-286-981B-16
; Sequence 16, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
```



PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 16  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-16

Query Match 55.4%; Score 1405.5; DB 4; Length 414;  
Best Local Similarity 67.7%; Pred. No. 8.9e-91;  
Matches 310; Conservative 45; Mismatches 46; Indels 57; Gaps 11;

37 TENEGATQVPTSSNRANESQAEOGPCKUDSERDKARKEVEEYVKTIAGESYAKSTKR 95  
1 TENEGSTQAATSSNMAK-----TEHKKAKQVVDYEITKMLRE--IQDLRR 44  
96 RHITVALVNLNINKNEYLNK--IVESTSESQOILMMESRSKYDEAVSKFEKSSSSS 153  
45 KHTQVALNINKLSAIKTYRLNVLNLEEKSKDELPE--SEIKAKUDAA--FEK----- 92  
154 SSDSTKEASDTAKPNKPTPEGEKVAAKKVVEA--KKAKDQKEDERNPTTYKTLE 212  
93 -----EKKDTLK-----PGEKVAAKKVVEAKKAAEDQKEDERNPTTYKTLE 138  
213 LEIAESDVEYKKAELVVKVANEPRDEOKIKQAEAVESKQAEATRLKIKTDREAEAE 272  
139 LEIAFDVKVAEAELELVKEAKESRNNGTIKQAEKESKKAETRLNITKDRKAAE 198  
273 EAKRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSP 323  
199 EAKRADAKLKEANVAATSDQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSS 258  
324 SLKPEKVAEAEKVEEAKKAAEDQKEDERNPTTYKTLELEIAESDVEYKKAELV 383  
259 SLKSGKVAEAEKVEEAKKAAEDQKEDERNPTTYKTLELEIAESDVKVKAELV 318  
384 KEBAKEPNEEKVKQAKAEVSKKAETRLKIKTDREAEAEAKKAAEDVKVEKPAE 443  
319 KEBAKEPNEEKIKQAKAKVESKKAETRLNITKDRKAAEAEAKKAAEDVKVEKPAE 378  
444 QOPAPAPKAKEKPAAPKPPENPAOPKAKEPADQAAE 481  
379 QOPAPATQPEK--PAPPEKPAEQPAEKTDQAAE 414

RESULT 14  
US-09-286-981B-10  
Sequence 10, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-136  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US/09/286, 981B  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-10

Query Match 55.2%; Score 1401.5; DB 4; Length 414;  
Best Local Similarity 66.6%; Pred. No. 1.7e-90;  
Matches 305; Conservative 48; Mismatches 48; Indels 57; Gaps 9;

37 TENEGATQVPTSSNRANESQAEOGPCKUDSERDKARKEVEEYVKTIAGESYAKSTKR 95

|||||  
1 TENEGSTQAATSSNMAK-----TEHKKAKQVVDYEITKMLRE--IQDLRR 44  
96 RHITVALVNLNINKNEYLNK--IVESTSESQOILMMESRSKYDEAVSKFEKSSSSS 153  
45 KHTQVALNINKLSAIKTYRLNVLNLEEKSKDELPE--SEIKAKUDAAFEKFKDQ----- 97  
154 SSDSTKEASDTAKPNKPTPEGEKVAAKKVVEA--KKAKDQKEDERNPTTYKTLE 212  
98 -----LKPGEKVAAKKVVEAKKAAEDQKEDERNPTTYKTLE 138  
213 LEIAESDVEYKKAELVVKVANEPRDEOKIKQAEAVESKQAEATRLKIKTDREAEAE 272  
139 LEIAFDVKVAEAELELVKEAKESRNNGTIKQAEKESKKAETRLNITKDRKAAE 198  
273 EAKRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSP 323  
199 EAKRADAKLKEANVAATSDQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSS 258  
324 SLKPEKVAEAEKVEEAKKAAEDQKEDERNPTTYKTLELEIAESDVEYKKAELV 383  
259 SLKSGKVAEAEKVEEAKKAAEDQKEDERNPTTYKTLELEIAESDVKVKAELV 318  
384 KEBAKEPNEEKVKQAKAEVSKKAETRLKIKTDREAEAEAKKAAEDVKVEKPAE 443  
319 KEBAKEPNEEKIKQAKAKVESKKAETRLNITKDRKAAEAEAKKAAEDVKVEKPAE 378  
444 QOPAPAPKAKEKPAAPKPPENPAOPKAKEPADQAAE 481  
379 QOPAPATQPEK--PAPPEKPAEQPAEKTDQAAE 414

RESULT 15  
US-09-286-981B-3  
Sequence 3, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-136  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US/09/286, 981B  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-3

Query Match 55.0%; Score 1395; DB 4; Length 431;  
Best Local Similarity 64.0%; Pred. No. 5.1e-90;  
Matches 304; Conservative 57; Mismatches 60; Indels 54; Gaps 9;

37 TENEGATQVPTSSNRANESQAEOGPCKUDSERDKARKEVEEYVKTIAGESYAKSTKR 96  
1 TEKEVTTTPVATSSNMKNSQ-----TEHKKAAEVDYEITNMI-----QDLKRX 44  
97 HTITVALVNLNINKNEYLNKIVESTSESQOILMMESRSKYDEAVSKFEKSSSSSSSD 156  
45 HTQVALNINKLSAIKTYRLNVLNLEEKSKDELTSKXKIDDAFQFQNDQ----- 97  
157 SSTKPEASDTAKPNKPTPEGEKVAAKKVVEA--KKAKDQKEDERNPTTYKTLEI 215  
98 -----LKPGEKVAAKKVVEAKKAAEDQKEDERNPTTYKTLEI 141  
216 AESDVEYKKAELVVKVANEPRDEOKIKQAEAVESKQAEATRLKIKTDREAEAEAEK 275  
142 AESDVEYKKAELVVEEAKSGNEEKIKQAKAEVSKKAETRLKIKTERKAAEAEK 201

```

Qy 276 RRAADAKQ--GKPKGRARQVPGELATPDKKENDAKSSDSSVGEETLPSLSIK----- 326
Db 202 RKAEEBEEVKNNLKKRTIRGAFGEPAIPDKKENDAKSSDSSSVKKS-SKPIIKSEKKVAE 260
Qy 327 PEKVAEAEKKEVEAEKKAEDQKEEDRRNYPNTYKTTLELEIAESDVEKKALELVKEE 386
Db 261 AEKVAEAEKKEVEAEKKAEDQKEEDRRNYPNTYKTTLELEIAESDVKKEAELELVKEE 320
Qy 387 AKEPNEEKVKQAKAEVSKAEATRLKIKITDRKKAEEAKRKAEEEDVKKEKPAEQPQ 446
Db 321 AKEPNEEKIKQAKAKVESKKAETRLKIKITDRKKA-EEAKRKAEEEDVKKEKPAEQPQ 379
Qy 447 PAPAPKAEKPAAPKPEENPAEQPKAEKPADQQAEEEDVARSSEENRLLTQQQPK 501
Db 380 PAPAP--KPAAPQPEKPAEQPKAEKPADQQAEEEDVARSSEENRLLTAPAK 431

```

Search completed: November 21, 2003, 13:35:10  
 Job time : 22.4457 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 7.55427 Seconds  
(without alignments)  
1008.166 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442

Perfect score: 889

Sequence: 1 IKTDREAEAEAKRADAKE.....AEEAKKAKAEEDKVEKPA 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/aa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/1/aa/5B.COMB.pep:\*

3: /cgn2\_6/ptodata/1/aa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/1/aa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/1/aa/PTUTS.COMB.pep:\*

6: /cgn2\_6/ptodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	446	US-09-286-981B-9	Sequence 9, Appl1
2	889	100.0	453	US-08-961-083-38	Sequence 38, Appl1
3	889	100.0	453	US-09-536-784-38	Sequence 38, Appl1
4	884	99.4	446	US-09-286-981B-6	Sequence 6, Appl1
5	803.5	90.4	414	US-09-286-981B-16	Sequence 16, Appl1
6	803.5	90.4	631	US-08-847-065-25	Sequence 25, Appl1
7	801.5	90.2	564	US-09-308-022-6	Sequence 6, Appl1
8	792	89.1	406	US-09-286-981B-18	Sequence 18, Appl1
9	792	89.1	425	US-09-286-981B-11	Sequence 11, Appl1
10	791.5	89.0	414	US-09-286-981B-10	Sequence 10, Appl1
11	784	88.2	428	US-09-286-981B-7	Sequence 7, Appl1
12	775.5	87.2	412	US-09-286-981B-17	Sequence 17, Appl1
13	770.5	86.7	424	US-09-286-981B-14	Sequence 14, Appl1
14	770.5	86.7	426	US-09-286-981B-12	Sequence 12, Appl1
15	769.5	86.6	425	US-09-286-981B-13	Sequence 13, Appl1
16	768	86.4	419	US-09-286-981B-15	Sequence 15, Appl1
17	717.5	80.7	413	US-09-286-981B-5	Sequence 5, Appl1
18	684.5	77.0	864	US-08-714-741-40	Sequence 40, Appl1
19	676.5	76.1	431	US-09-286-981B-3	Sequence 3, Appl1
20	669	75.3	588	US-08-714-741-42	Sequence 42, Appl1
21	557.5	62.7	219	US-09-286-981B-8	Sequence 8, Appl1
22	542	61.0	114	US-09-286-981B-19	Sequence 19, Appl1
23	521	58.6	142	US-08-847-065-21	Sequence 21, Appl1
24	517	58.2	251	US-09-286-981B-4	Sequence 4, Appl1
25	487	54.8	103	US-09-286-981B-1	Sequence 1, Appl1
26	461.5	51.9	605	US-08-714-741-46	Sequence 46, Appl1
27	451	50.7	103	US-09-286-981B-38	Sequence 38, Appl1

28	447.5	50.3	1231	US-08-714-741-41	Sequence 41, Appl1
29	394.5	44.4	623	US-08-714-741-47	Sequence 47, Appl1
30	375.5	42.2	8991	US-08-714-741-32	Sequence 32, Appl1
31	306	34.4	110	US-08-961-083-102	Sequence 102, Appl1
32	306	34.4	110	US-09-536-784-102	Sequence 102, Appl1
33	237.5	26.7	55	US-09-286-981B-37	Sequence 37, Appl1
34	182.5	20.5	1507	US-08-929-329-5	Sequence 5, Appl1
35	174	19.6	468	US-09-328-352-6321	Sequence 6321, Appl1
36	159.5	17.9	611	US-09-216-393B-81	Sequence 81, Appl1
37	158.5	17.8	288	US-08-312-949-4	Sequence 4, Appl1
38	158.5	17.8	288	US-08-446-201-4	Sequence 4, Appl1
39	158.5	17.8	619	US-08-465-746-2	Sequence 2, Appl1
40	158.5	17.8	619	US-08-214-164-2	Sequence 2, Appl1
41	158.5	17.8	619	US-08-467-852A-3	Sequence 3, Appl1
42	158.5	17.8	619	US-08-246-636-2	Sequence 2, Appl1
43	158.5	17.8	619	US-08-247-491A-3	Sequence 3, Appl1
44	158.5	17.8	619	US-08-319-795-2	Sequence 2, Appl1
45	158.5	17.8	619	US-08-468-985-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-09-286-981B-9  
Sequence 9, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 9  
LENGTH: 446  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-9

Query Match 100.0%; Score 889; DB 4; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.2e-65;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKEOGKPKGRARGVGELATDPDKKNDKSSSVGEETLPS 60  
228 IKTDREAEAEAKRADAKEOGKPKGRARGVGELATDPDKKNDKSSSVGEETLPS 287

Db 61 PSLKPKKVAEAKKVEAKKKAEDOKEDRRVYPTNTYKTLLELAESDVEYKKALEL 120  
288 PSLKPKKVAEAKKVEAKKKAEDOKEDRRVYPTNTYKTLLELAESDVEYKKALEL 347

QY 121 VKEAEAPNEEKVKAKAEVESKKAATRLKIKTKRKAEEAKRKAEDVKVEKPA 180  
348 VKEAEAPNEEKVKAKAEVESKKAATRLKIKTKRKAEEAKRKAEDVKVEKPA 407

RESULT 2  
US-08-961-083-38  
Sequence 38, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-38

Query Match 100.0%; Score 889; DB 3; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.3e-65;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60  
25 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 84  
61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 120  
85 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 144  
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEBAKRAAEEDVKKEKPA 180  
145 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEBAKRAAEEDVKKEKPA 204

RESULT 3  
US-09-536-784-38  
Sequence 38, Application US/09536784  
Patent No. 6573082

GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA

ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-09-536-784-38

Query Match 100.0%; Score 889; DB 4; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.3e-65;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60  
25 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 84  
61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 120  
85 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 144  
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEBAKRAAEEDVKKEKPA 180  
145 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEBAKRAAEEDVKKEKPA 204

RESULT 4  
US-09-286-981B-6  
Sequence 6, Application US/09286981B  
Patent No. 6503511

GENERAL INFORMATION:  
APPLICANT: Wizenann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 446  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-6

Query Match 99.4%; Score 884; DB 4; Length 446;  
Best Local Similarity 99.4%; Pred. No. 3.2e-65;  
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60  
228 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 287  
61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 120  
288 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKKALEL 347  
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEBAKRAAEEDVKKEKPA 180

Db 348 VKEAKEPRNEKVKQAAVESEKAEATRLKIKTRDKKAEBAKRAAEEDKVKRPA 407

RESULT 5  
US-09-286-981B-16  
Sequence 16, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 16  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-16

Query Match 90.4%; Score 803.5; DB 4; Length 414;  
Best Local Similarity 86.8%; Pred. No. 1.2e-58;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTRDEAEAEBAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51  
Db 189 IKTRDKKAEBAKRRADAKLKEANVATSDQKPKGRAGVPGELATPDKKENDAKSSDS 248  
QY 52 SVGEETLPSSLSKPKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLDEIAESDV 111  
Db 249 SVGEETLPSSLSKPKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLDEIAESDV 308  
QY 112 EVKKAELVKEBAKPRNEKVKQAAVESEKAEATRLKIKTRDKKAEBAKRAAE 171  
Db 309 KVKAELELVKEBAKPRDEKIKQAKAKVESKAEATRLKIKTRDKKAEBAKRAAE 368  
QY 172 EDKVKRPA 180  
Db 369 EDKVKRPA 377

RESULT 6  
US-08-847-065-25  
Sequence 25, Application US/08847065  
Patent No. 6245335  
GENERAL INFORMATION:  
APPLICANT: Masure, H. Robert  
APPLICANT: Rosenow, Carsten I.  
APPLICANT: Tuomanen, Elaine  
APPLICANT: Witzmann, Theresa M.  
TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/847,065

FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-158  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 631 amino acids  
TYPE: amino acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-847-065-25

Query Match 90.4%; Score 803.5; DB 3; Length 631;  
Best Local Similarity 86.8%; Pred. No. 2e-58;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTRDEAEAEBAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51  
Db 155 IKTRDKKAEBAKRRADAKLKEANVATSDQKPKGRAGVPGELATPDKKENDAKSSDS 214  
QY 52 SVGEETLPSSLSKPKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLDEIAESDV 111  
Db 215 SVGEETLPSSLSKPKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLDEIAESDV 274  
QY 112 EVKKAELVKEBAKPRNEKVKQAAVESEKAEATRLKIKTRDKKAEBAKRAAE 171  
Db 275 KVKAELELVKEBAKPRDEKIKQAKAKVESKAEATRLKIKTRDKKAEBAKRAAE 334  
QY 172 EDKVKRPA 180  
Db 335 EDKVKRPA 343

RESULT 7  
US-09-308-022-6  
Sequence 6, Application US/09308022  
Patent No. 6291654  
GENERAL INFORMATION:  
APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.  
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.  
STREET: 119 No. 6291654th Fourth Street, Suite 203  
CITY: Minneapolis  
STATE: Minnesota  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/308,022  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/20586  
FILING DATE: 12-NOV-1997  
APPLICATION NUMBER: 60/029,444  
FILING DATE: 12-NOV-1996  
APPLICATION NUMBER: 60/038,086  
FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368  
FILING DATE: 19-SEP-1997  
APPLICATION NUMBER: 60/062,473  
FILING DATE: 16-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: MUETING, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00430101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-308-022-6

Query Match 90.2%; Score 801.5; DB 3; Length 564;  
Best Local Similarity 86.2%; Pred. No. 2.5e-58;  
Matches 163; Conservative 13; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDS 51  
DB 226 IKTDRBEAEAEKRRADAKLKEANVATSDQKPKGRARGVPGELATPDKKENDAKSSDS 285  
QY 52 SVGEETLPSPLPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDV 111  
DB 286 SVGEETLPSPLSSSKGKVAEAEKKVEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 345  
QY 112 EVKKAELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRAE 171  
DB 346 KVEKAELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRAE 405  
QY 172 EDKVEKPA 180  
DB 406 EDKVEKPA 414

RESULT 8  
US-09-286-981B-18  
Sequence 18, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 18  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-18

Query Match 89.1%; Score 792; DB 4; Length 406;  
Best Local Similarity 90.2%; Pred. No. 1e-57;  
Matches 166; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAKGCG---KPKGRARGVPGELATPDKKENDAKSSDSVCEE 56  
DB 188 IKTDRBEAEAEKRRADAKGCG---KPKGRARGVPGELATPDKKENDAKSSDSVCEE 247  
QY 57 TLPSPSLPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDVAKKA 116

DB 248 TLPSPSLPEKKVAEAEKKVEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDVAKKA 307  
QY 117 ELELVKEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRAEEDKVK 176  
DB 308 ELELVKEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRAEEDKVK 367  
QY 177 EKPA 180  
DB 368 EKPA 371

RESULT 9  
US-09-286-981B-11  
Sequence 11, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 11  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-11

Query Match 89.1%; Score 792; DB 4; Length 425;  
Best Local Similarity 85.4%; Pred. No. 1.1e-57;  
Matches 164; Conservative 8; Mismatches 8; Indels 12; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAK-----EOKPKGRARGVPGELATPDKKENDAKS 48  
DB 188 IKTDRBEAEAEKRRADAKLKEANVATSDQKPKGRARGVPGELATPDKKENDAKS 247  
QY 49 SDSSVGEETLPSPLPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAE 108  
DB 248 SDSSVGEETLPSPLPEKKVAEAEKKVEAEKKAEDQKEEDRRNYPNTYKTLLEIAE 307  
QY 109 SDVEKKAELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRK 168  
DB 308 SDVKESELELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKITDRKAAEEAKRK 367  
QY 169 AAEDKVEKPA 180  
DB 368 AAEDKVEKPA 379

RESULT 10  
US-09-286-981B-10  
Sequence 10, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 414  
TYPE: PRT

ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-10

Query Match 89.0%; Score 791.5; DB 4; Length 414;  
Best Local Similarity 85.7%; Pred. No. 1.2e-57;  
Matches 162; Conservative 12; Mismatches 6; Indels 9; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 189 IKTDREAEAEAKRRADAKKEANVATSDQKPKGRAGVPGELATPDKKENDAKSSDS 248  
QY 52 SVGETLPSPLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDSV 111  
DB 249 SVGETLPSSSLKSGKVAEAEKKVEAEKKAEDQKEDRRNPTNTYKTLLEIASDV 308  
QY 112 EVKKAELVKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAE 171  
DB 309 KVKAELELVKEAEKPEKDEKIKQAAYESKKAETRLKIKTDKKAEEBAKRAKAE 368  
QY 172 EDKVKEKPA 180  
DB 369 EDKVKEKPA 377

RESULT 11  
US-09-286-981B-7

Sequence 7, Application US/09286981B  
Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

CURRENT FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 7

LENGTH: 428

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-7

Query Match 88.2%; Score 784; DB 4; Length 428;  
Best Local Similarity 90.6%; Pred. No. 5e-57;

Matches 163; Conservative 4; Mismatches 7; Indels 6; Gaps 1;

QY 1 IKTDREAEAEAKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 60  
DB 209 IKTDREAEAEAE-----AKVDEPKKRTKRGVLGEPATPDKKENDAKSSDSVGEETLPS 262  
QY 61 PSLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEVKAELEL 120  
DB 263 PSLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEVKAELEL 322  
QY 121 VKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAEEDVKKEPA 180  
DB 323 VKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAEEDVKKEPA 362

RESULT 12  
US-09-286-981B-17

Sequence 17, Application US/09286981B  
Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 17

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 57  
DB 191 IKTDREAEAEAKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 250

Query Match 87.2%; Score 775.5; DB 4; Length 412;  
Best Local Similarity 89.6%; Pred. No. 2.4e-56;  
Matches 164; Conservative 4; Mismatches 12; Indels 3; Gaps 1;

QY 1 IKTDREAEAEAKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 57  
DB 191 IKTDREAEAEAKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 250  
QY 58 PSLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEVKAE 117  
DB 251 PSLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEVKAE 310  
QY 118 LEVKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAEEDVKKE 177  
DB 311 LEVKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAEEDVKKE 370  
QY 178 KPA 180  
DB 371 KPA 373

RESULT 13

US-09-286-981B-14

Sequence 14, Application US/09286981B

Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

CURRENT FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 14

LENGTH: 424

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-14

Query Match 86.7%; Score 770.5; DB 4; Length 424;  
Best Local Similarity 87.4%; Pred. No. 6.4e-56;  
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 190 IKTDREAEAEAKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDS 248  
QY 52 SVGETLPSPLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDSV 111  
DB 249 SVGETLPSPLKPEKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDV 308  
QY 112 EVKKAELVKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAE 170  
DB 309 EVKKAELVKEAEKPEKNEKVKQAAYESKKAETRLKIKTDKKAEEBAKRAKAE 368  
QY 171 EDKVKEKPA 180

Db 369 EEDKVEKPA 378

## RESULT 14

US-09-286-981B-12  
; Sequence 12, Application US/09286981B  
; Patent No. 6503511

## GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 12

LENGTH: 426

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-12

Query Match 86.7%; Score 770.5; DB 4; Length 426;

Best Local Similarity 87.4%; Pred. No. 6.4e-56;

Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDS 51

Db 190 IKTDREKA-EEAKRRADAKLEANVATSEODSKRRAREVLGELATPDKKENDAKSSDS 248

QY 52 SVGEETLPSPSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNYPNTYKTLLEIAESDV 111

Db 249 SVGEETLPSPSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNYPNTYKTLLEIAESDV 308

QY 112 EVKAELELVKEAEKPEPNEEKVQAKAEVSKKAATRLLEKIKTDREKA-EEBAKRAA 170

Db 309 EVKAELELVKEAEKPEPNEEKVQAKAEVSKKAATRLLEKIKTDREKA-EEBAKRAA 368

QY 171 EEDKVEKPA 180

Db 369 EEDKVEKPA 378

## RESULT 15

US-09-286-981B-13  
; Sequence 13, Application US/09286981B  
; Patent No. 6503511

## GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 13

LENGTH: 425

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-13

Query Match 86.6%; Score 769.5; DB 4; Length 425;

Best Local Similarity 87.4%; Pred. No. 7.7e-56;

Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDS 51

Db 190 IKTDREKA-EEAKRRADAKLEANVATSEODSKRRAREVLGELATPDKKENDAKSSDS 248

QY 52 SVGEETLPSPSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNYPNTYKTLLEIAESDV 111

Db 249 SVGEETLPSPSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNYPNTYKTLLEIAESDV 308

QY 112 EVKAELELVKEAEKPEPNEEKVQAKAEVSKKAATRLLEKIKTDREKA-EEBAKRAA 170

Db 309 EVKAELELVKEAEKPEPNEEKVQAKAEVSKKAATRLLEKIKTDREKA-EEBAKRAA 368

QY 171 EEDKVEKPA 180

Db 369 EEDKVEKPA 378

Search completed: November 21, 2003, 13:35:11

Job time : 8.55427 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:33:52 ; Search time 39.1939 Seconds  
(without alignments)  
2380.162 Million cell updates/sec

Title: US-09-298-523C-13  
Perfect score: 2537  
Sequence: 1 MFASKSRKHYIRKFSVG.....NRLTQQQPKTEKAPQSTP 511

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues  
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	2537	100.0	511	11	US-09-298-523B-13	Sequence 13, Appl
2	2526	99.6	513	11	US-09-298-523B-12	Sequence 12, Appl
3	2516	99.2	693	11	US-09-769-787-185	Sequence 185, App
4	2350.5	92.6	655	11	US-09-056-019-2	Sequence 2, Appl1
5	2252.5	88.8	460	11	US-09-056-019-38	Sequence 38, Appl
6	2235.5	88.1	459	11	US-09-056-019-39	Sequence 39, Appl
7	2194.5	86.5	446	12	US-10-254-995-9	Sequence 9, Appl1
8	2177.5	85.8	446	12	US-10-254-995-6	Sequence 6, Appl1
9	2105	83.0	451	11	US-09-298-523B-67	Sequence 67, Appl
10	2101.5	82.8	428	11	US-09-056-019-24	Sequence 24, Appl
11	1979.5	78.0	406	11	US-09-056-019-1	Sequence 1, Appl1
12	1761.5	69.4	487	11	US-09-298-523B-66	Sequence 66, Appl
13	1745.5	68.8	487	11	US-09-298-523B-66	Sequence 66, Appl
14	1718	67.7	581	11	US-09-298-523B-56	Sequence 56, Appl
15	1715	67.6	483	11	US-09-298-523B-10	Sequence 10, Appl

16	1700	67.0	481	11	US-09-298-523B-6	Sequence 6, Appl1
17	1679.5	66.2	496	11	US-09-298-523B-4	Sequence 4, Appl1
18	1569.5	61.9	663	11	US-09-298-523B-58	Sequence 58, Appl
19	1569.5	61.9	663	11	US-09-056-019-8	Sequence 8, Appl1
20	1569.5	61.9	663	12	US-09-969-748C-14	Sequence 14, Appl
21	1569.5	61.9	663	12	US-09-949-039-104	Sequence 31, Appl
22	1568.5	61.8	419	11	US-09-056-019-31	Sequence 31, Appl
23	1559.5	61.5	478	11	US-09-298-523B-65	Sequence 65, Appl
24	1534	60.5	437	11	US-09-298-523B-64	Sequence 64, Appl
25	1526	60.1	444	11	US-09-298-523B-5	Sequence 5, Appl1
26	1512.5	59.6	406	12	US-10-254-995-18	Sequence 18, Appl
27	1506	59.4	631	12	US-09-829-382-25	Sequence 25, Appl
28	1504.5	59.3	439	11	US-09-056-019-28	Sequence 28, Appl
29	1497.5	59.0	437	11	US-09-056-019-32	Sequence 32, Appl
30	1496.5	59.0	437	11	US-09-056-019-29	Sequence 29, Appl
31	1494	58.9	439	11	US-09-056-019-30	Sequence 30, Appl
32	1465.5	57.8	439	11	US-09-056-019-37	Sequence 37, Appl
33	1465.5	57.8	433	11	US-09-056-019-33	Sequence 33, Appl
34	1455.5	57.4	426	12	US-10-254-995-12	Sequence 12, Appl
35	1447.5	57.1	425	12	US-10-254-995-13	Sequence 13, Appl
36	1446.5	57.0	424	12	US-10-254-995-14	Sequence 14, Appl
37	1440.5	56.8	428	12	US-10-254-995-7	Sequence 7, Appl1
38	1436.5	56.6	425	12	US-09-056-019-36	Sequence 36, Appl
39	1432	56.4	425	12	US-10-254-995-11	Sequence 11, Appl
40	1412.5	55.7	412	12	US-10-254-995-17	Sequence 17, Appl
41	1410.5	55.6	412	12	US-10-254-995-15	Sequence 15, Appl
42	1408.5	55.5	413	11	US-09-056-019-35	Sequence 35, Appl
43	1405.5	55.4	414	12	US-10-254-995-16	Sequence 16, Appl
44	1401.5	55.2	414	12	US-10-254-995-10	Sequence 10, Appl
45	1399	55.1	711	11	US-09-298-523B-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-09-298-523B-13  
; Sequence 13, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OR INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298, 523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
; US-09-298-523B-13

Query Match 100.0%; Score 2537; DB 11; Length 511;  
Best Local Similarity 100.0%; Pred. No. 3.9e-121;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MFASKSRKHYIRKFSVGVASVWASVWVHATENEGATOVPTSSNRANESQAEQG 60  
DB 1 MFASKSRKHYIRKFSVGVASVWASVWVHATENEGATOVPTSSNRANESQAEQG 60  
QY EQPKTIDSEBDKARKVEEYVKKIIVGSAKSTYKHTTTVALVNEIINIKNEYLNKIVE 120  
DB EQPKTIDSEBDKARKVEEYVKKIIVGSAKSTYKHTTTVALVNEIINIKNEYLNKIVE 120  
QY STEESQLOILMMSRKVDVAVSKFEKDSSSSSSTKPEASDPAKPKPEPEKVA 180  
DB STEESQLOILMMSRKVDVAVSKFEKDSSSSSSTKPEASDPAKPKPEPEKVA 180  
QY EAKKVEEAKKADQKEEDRRNPPTITTYKLELEIAESDVEVKAKLELVKANPRDE 240  
DB EAKKVEEAKKADQKEEDRRNPPTITTYKLELEIAESDVEVKAKLELVKANPRDE 240

Qy 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300  
Db 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300  
Qy 301 PDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPTNT 360  
Db 301 PDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPTNT 360  
Qy 361 YKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKTDR 420  
Db 361 YKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKTDR 420  
Qy 421 KKAEEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQAE 480  
Db 421 KKAEEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQAE 480  
Qy 481 EDVARSSEEEYNRLTQOOPPKTEKPAQPSPT 511  
Db 481 EDVARSSEEEYNRLTQOOPPKTEKPAQPSPT 511

## RESULT 2

US-09-298-523B-12  
; Sequence 12, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-12

Query Match 99.6%; Score 256; DB 11; Length 513;  
Best Local Similarity 99.6%; Pred. No. 1.4e-120;  
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
Db 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
Qy 61 EOPKRLDSEBDKARKEVEEYVKKIIGESYAKSTKKRHITVALVNLNINIKNEYLINKIV 120  
Db 61 EOPKRLDSEBDKARKEVEEYVKKIIGESYAKSTKKRHITVALVNLNINIKNEYLINKIV 120  
Qy 121 STSESQLOILMMESRSKVDKAVSKPEKSSSSSSSTKPEASDTAKPNKPTPEGEKVA 180  
Db 121 STSESQLOILMMESRSKVDKAVSKPEKSSSSSSSTKPEASDTAKPNKPTPEGEKVA 180  
Qy 181 EAKKVEBAKAKQCKEDRRNYPTITYKTLLEIAESDVEYKKALELVKVKANEPDE 240  
Db 181 EAKKVEBAKAKQCKEDRRNYPTITYKTLLEIAESDVEYKKALELVKVKANEPDE 240  
Qy 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300  
Db 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300  
Qy 301 PDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPTNT 360  
Db 301 PDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPTNT 360  
Qy 361 YKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKTDR 420  
Db 361 YKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKTDR 420

Qy 421 KKAEEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQAE 480  
Db 421 KKAEEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQAE 480  
Qy 481 EDVARSSEEEYNRLTQOOPPKTEKPAQPSPT 511  
Db 481 EDVARSSEEEYNRLTQOOPPKTEKPAQPSPT 511

## RESULT 3

US-09-769-787-185  
; Sequence 185, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: GB 9616337.1  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 185  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-185

Query Match 99.2%; Score 2516; DB 11; Length 693;  
Best Local Similarity 99.6%; Pred. No. 6.1e-120;  
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 59  
Db 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60  
Qy 60 GEOPKRLDSEBDKARKEVEEYVKKIIGESYAKSTKKRHITVALVNLNINIKNEYLINKIV 119  
Db 60 GEOPKRLDSEBDKARKEVEEYVKKIIGESYAKSTKKRHITVALVNLNINIKNEYLINKIV 120  
Qy 120 ESTSESQLOILMMESRSKVDKAVSKPEKSSSSSSSTKPEASDTAKPNKPTPEGEKVA 179  
Db 120 ESTSESQLOILMMESRSKVDKAVSKPEKSSSSSSSTKPEASDTAKPNKPTPEGEKVA 180  
Qy 180 AEAKKYVEA-KKAKQCKEDRRNYPTITYKTLLEIAESDVEYKKALELVKVKANEPDE 238  
Db 180 AEAKKYVEA-KKAKQCKEDRRNYPTITYKTLLEIAESDVEYKKALELVKVKANEPDE 240  
Qy 239 DEQKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGEL 298  
Db 239 DEQKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGEL 300  
Qy 299 ATPDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPT 358  
Db 299 ATPDKKENDAKSSDSVGEETLPSPLKPKKVAEAKVEAKKAEADQCKEDRRNYPT 360  
Qy 359 NTYKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKT 418  
Db 359 NTYKTLLEIAESDVEYKKALELVKEAKEPRNEEKVQAKAEVSKAEATRLKIKT 420  
Qy 419 DRKAEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQ 478  
Db 419 DRKAEBAKRAAEEDKVEKPAEQOPAPAPKAEKPAKPEKPAEQPKAEKPADQ 480  
Qy 479 AEEDYARSSEEEYNRLTQOOPPKTEKPAQPSPT 511  
Db 479 AEEDYARSSEEEYNRLTQOOPPKTEKPAQPSPT 513

```
RESULT 4
US-09-056-019-2
; Sequence 2, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT FILING DATE: 1998-04-07
; CURRENT APPLICATION NUMBER: US/09/056,019A
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-2

Query Match      92.6%; Score 2350.5; DB 11; Length 655;
Best Local Similarity 99.8%; Pred. No. 1,4e-111;
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKRRH 97
DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKRRH 60
QY 98 TITVALVNELNINKIYINKIYESTSESQIILMESRSKYDEAVSKPEKSSSSSSSDS 157
DB 61 TITVALVNELNINKIYINKIYESTSESQIILMESRSKYDEAVSKPEKSSSSSSSDS 120
QY 158 STKPSASTAKPNKTEPGEKVAEAKKVEEA-KKAKOKEEDRNNTYITIKTLELEIA 216
DB 121 STKPSASTAKPNKTEPGEKVAEAKKVEEA-KKAKOKEEDRNNTYITIKTLELEIA 180
QY 217 ESDVEVKAELVVKANEPDEOKIKQAEAVESKQAEATRLKIKTDRBEAEBAKR 276
DB 181 ESDVEVKAELVVKANEPDEOKIKQAEAVESKQAEATRLKIKTDRBEAEBAKR 240
QY 277 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKPEKTVAEAK 336
DB 241 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKPEKTVAEAK 300
QY 337 KVEAKKKAEADQKEDRNNTYITIKTLEIASDVVKAELELVKEAKKEPRNEKV 396
DB 301 KVEAKKKAEADQKEDRNNTYITIKTLEIASDVVKAELELVKEAKKEPRNEKV 360
QY 397 KQAAVEESKVAETRLKIKTDRKKAEBEAKRKAABEDKVEKPAEQOPAPAKAEP 456
DB 361 KQAAVEESKVAETRLKIKTDRKKAEBEAKRKAABEDKVEKPAEQOPAPAKAEP 420
QY 457 AAPAKPENPAEQPAEKPADQQAEDVARRSEEEYNRLTQQQPKTEKPAQSTP 511
DB 421 AAPAKPENPAEQPAEKPADQQAEDVARRSEEEYNRLTQQQPKTEKPAQSTP 475

RESULT 5
US-09-056-019-38
; Sequence 38, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
```

```
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match      88.8%; Score 2252.5; DB 11; Length 460;
Best Local Similarity 99.6%; Pred. No. 8.6e-107;
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWGSVVAHNEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVK 83
DB 2 IVASLWGSVVAHNEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVK 61
QY 84 IVGESYAKSTKRRHTITVALVNELNINKIYINKIYESTSESQIILMESRSKYDEAVS 143
DB 62 IVGESYAKSTKRRHTITVALVNELNINKIYINKIYESTSESQIILMESRSKYDEAVS 121
QY 144 KFEKSSSSSSSDSTKPEASDTAKPKPTPEPGKVAEAKKVEEA-KKAKDOKEEDRN 202
DB 122 KFEKSSSSSSSDSTKPEASDTAKPKPTPEPGKVAEAKKVEEA-KKAKDOKEEDRN 181
QY 203 YPTITTYITLELEIESDVEVKAELVVKANEPDEOKIKQAEAVESKQAEATRLK 262
DB 182 YPTITTYITLELEIESDVEVKAELVVKANEPDEOKIKQAEAVESKQAEATRLK 241
QY 263 IKTDREAEBAEKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 322
DB 242 IKTDREAEBAEKRRADAKQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 301
QY 323 PSLKPEKVAEAKKVEEAKKKAADQKEDRNNTYITIKTLELEIESDVEVKAEL 382
DB 302 PSLKPEKVAEAKKVEEAKKKAADQKEDRNNTYITIKTLELEIESDVEVKAEL 361
QY 383 VKERAKPRNEKVQKAAVESKKAATRLKIKTDRKKAEBEAKRKAABEDVKEKPA 442
DB 362 VKERAKPRNEKVQKAAVESKKAATRLKIKTDRKKAEBEAKRKAABEDVKEKPA 421
QY 443 EQQAPAPAPAKKAPAKPENPAEQPAEKPADQQAEE 481
DB 422 EQQAPAPAPAKKAPAKPENPAEQPAEKPADQQAEE 460

RESULT 6
US-09-056-019-39
; Sequence 39, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-39
```

Query Match 88.1%; Score 2235.5; DB 11; Length 459;  
Best Local Similarity 98.7%; Pred. No. 6.2e-106;  
Matches 453; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

OY 24 VVASLWMSVYATNEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYK 83  
DB 1 IVASLWMSVYATNEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYK 60

OY 84 IVGESYAKSTKKRHTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVS 143  
DB 61 IVGESYAKSTKKRHTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVS 120

OY 144 KPEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAEKKKVEEA-KKAKDQKEEDRN 202  
DB 121 KPEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAEKKKVEEA-KKAKDQKEEDRN 180

OY 203 YPTITVKTLLEIAESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLK 262  
DB 181 YPTITVKTLLEIAESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLK 240

OY 263 IKTDREAEAEKRRADAKQCKPKRPGVPGELATPDKKENDAKSSDSSVGEETLPS 322  
DB 241 IKTDREAEAEKRRADAKQCKPKRPGVPGELATPDKKENDAKSSDSSVGEETLPS 300

OY 323 PSLEKEKVAEAEKVEEAKKAEKEDKEDRNPTNTYKTLLEIAESDVEVKAELEL 382  
DB 301 PSLEKEKVAEAEKVEEAKKAEKEDKEDRNPTNTYKTLLEIAESDVEVKAELEL 360

OY 383 VKEAEKERNNEKVAEAEVSKKAEATRLKIKTDREAEAEKRRADAKQCKPKR 442  
DB 361 VKEAEKERNNEKVAEAEVSKKAEATRLKIKTDREAEAEKRRADAKQCKPKR 420

OY 443 EQOPAPAPKAEKPAAPKPEBPAPQPAEKPADQAE 481  
DB 421 EQOPAPAPKAEKPAAPKPEBPAPQPAEKPADQAE 459

RESULT 7  
US-10-254-995-9  
; Sequence 9, Application US/10254995  
; Publication No. US20030138447A1  
; GENERAL INFORMATION:  
; APPLICANT: Witzemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/10/254,995  
; CURRENT FILING DATE: 2002-09-25  
; PRIOR APPLICATION NUMBER: US/09/286,981  
; PRIOR FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-254-995-9

Query Match 86.5%; Score 2194.5; DB 12; Length 446;  
Best Local Similarity 99.8%; Pred. No. 7.2e-104;  
Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 37 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKTI VGESYAKSTKKR 96  
DB 1 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKTI VGESYAKSTKKR 60

OY 97 HTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVSKFEKSSSSSSSD 156  
DB 61 HTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVSKFEKSSSSSSSD 120

OY 157 SSTPEASDTAKPNKPTPEGKVAEAKKVEEA-KKAKDQKEEDRNPTITVKTLLEI 215  
DB 121 SSTPEASDTAKPNKPTPEGKVAEAKKVEEA-KKAKDQKEEDRNPTITVKTLLEI 180

OY 216 AESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 275  
DB 181 AESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 240

OY 276 RRADAKQCKPKRAGKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 335  
DB 241 RRADAKQCKPKRAGKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 300

OY 336 KKVAEAKKAAEDKEDRNPTNTYKTLLEIAESDVEVKAELELVKVEAEKERNNEK 395  
DB 301 KKVAEAKKAAEDKEDRNPTNTYKTLLEIAESDVEVKAELELVKVEAEKERNNEK 360

OY 396 VKQAEAVESKKAATRLKIKTDREAEAEKRRADAKQCKPKRPGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 455  
DB 361 VKQAEAVESKKAATRLKIKTDREAEAEKRRADAKQCKPKRPGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 420

OY 456 PAPAPKPEBPAPQPAEKPADQAE 481  
DB 421 PAPAPKPEBPAPQPAEKPADQAE 446

RESULT 8  
US-10-254-995-6  
; Sequence 6, Application US/10254995  
; Publication No. US20030138447A1  
; GENERAL INFORMATION:  
; APPLICANT: Witzemann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/10/254,995  
; CURRENT FILING DATE: 2002-09-25  
; PRIOR APPLICATION NUMBER: US/09/286,981  
; PRIOR FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-254-995-6

Query Match 85.8%; Score 2177.5; DB 12; Length 446;  
Best Local Similarity 98.9%; Pred. No. 5.2e-103;  
Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 37 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKTI VGESYAKSTKKR 96  
DB 1 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKTI VGESYAKSTKKR 60

OY 97 HTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVSKFEKSSSSSSSD 156  
DB 61 HTITVALVNLNINKNEYLNKIVESTSESQOILMMSRSKYDAVSKFEKSSSSSSSD 120

OY 157 SSTPEASDTAKPNKPTPEGKVAEAKKVEEA-KKAKDQKEEDRNPTITVKTLLEI 215  
DB 121 SSTPEASDTAKPNKPTPEGKVAEAKKVEEA-KKAKDQKEEDRNPTITVKTLLEI 180

OY 216 AESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 275  
DB 181 AESDVEVKAELELVKVKANEPREDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 240

OY 276 RRADAKQCKPKRAGKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 335  
DB 241 RRADAKQCKPKRAGKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 300

Qy	336	KVVEAKKKAADQEKEDERNYPTNYKTLELELASDVEKKAELTVBEAKPENNEK	395
Db	301	KVVEAKKKAADQEKEDERNYPTNYKTLELELASDVEKKAELTVBEAKPENNEK	360
Qy	396	VKQAAEYVESKKAELTRLEKITDRKAAEBAKRAEEDYKERAEOPOPAPAKAEK	455
Db	361	VKQAAEYVESKKAELTRLEKITDRKAAEBAKRAEEDYKERAEOPOPAPAKAEK	420
Qy	456	PAPAKPENPAEQVQAEXPADQAAE	481
Db	421	PAPAKPENPAEQVQAEXPADQAAE	446

## RESULT 9

US-09-298-523B-67  
 : Sequence 67, Application US/09298523B  
 : Publication No. US20030059438A1  
 : GENERAL INFORMATION:  
 : APPLICANT: BRILES et al.  
 : TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
 : TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
 : FILE REFERENCE: 454312-3140  
 : CURRENT APPLICATION NUMBER: US/09/298,523B  
 : CURRENT FILING DATE: 1999-04-23  
 : NUMBER OF SEQ ID NOS: 78  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 67  
 : LENGTH: 451  
 : TYPE: PRT  
 : ORGANISM: Streptococcus pneumoniae  
 : US-09-298-523B-67

Query Match	83.0%;	Score 2105;	DB 11;	Length 451;
Best Local Similarity	84.6%;	Pred. No. 2.4e-99;		
Matches 439; Conservative	1;	Mismatches	1;	Indels 78; Gaps 4;

QY	I	MPASKESEKHYSTIKFSYG-ASVYVASLWMGSVYHATENEGATQVPTSSNPNANESQAOE	59
	1	MPASKESEKHYVSIIRKFSVGAVASVVASLWMGSVYHATENEGATQVPTSSNPNANESQAOE	60
QY	60	GEOPKRLDSEBDKARKEVEEYVKTI VGSYASTKRRHITVALVNELNNINNEYLNKTV	119
Db	61	GEOPKRLDSEBDKARKEVEEYVKTI VGSYASTKRRHITVALVNELNNINNEYLNKTV	120
QY	120	ESTSESOLOIMMESRSKDEAVSKFEKDSSSSSSDSSTKPEASDTAKPNKTEGEKY	179
Db	121	ESTSESOQILIM-----	133
QY	180	AEAKKVEEAKKADQKEEDRRNYPTITYKTILELEIASDVEYKAELELVYKANEPD	239
Db	134	-----NYPTITYKTILELEIASDVEYKAELELVYKANEPD	171
QY	240	EOKIKOAEAEVSKOAEATRLKKITIDRBEAEEAKRRADAKOQKPKGRARGVGELA	299
Db	172	EOKIKOAEAEVSKOAEATRLKKITIDRBEAEEAKRRADAKOQKPKGRARGVGELA	231
QY	300	TPDKKENDAKSSDSVGEBTLPSBLKJBEKUYAEAEKYVEAKKKAEDQKEDRRNYPTN	359
Db	232	TPDKKENDAKSSDSVGEBTLPSBLKJBEKUYAEAEKYVEAKKKAEDQKEDRRNYPTN	291
QY	360	TYKTILELEIASDVEYKAELELVYKEAKPENNEKYVQO-----RAEVESKKAENR	412
Db	292	TYKTILELE-ASSDVEYKAELELVYKEAKPENNEKYVQOAKAEVSKKAEEVSKKAEBATR	350
QY	413	LEKIKTDKKAEEAKRKAABEDKYKEXKPAEOPQAPAPKAEXKPAKPENKPAEOPKAE	472
Db	351	LEKIKTDKKAEEAKRKAABEDKYKEXKPAEOPQAPAPKAEXKPAKPENKPAEOPKAE	410
QY	473	KPADQOAEEDYARSRSEEFYNNLTQOQPKTEKPADPSTP	511
Db	411	KPADQOAEEDYARSRSEEFYNNLTQOQPKTEKPADPSTP	449

```

RESULT 10
US-09-056-019-24
; Sequence 24, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koehnig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOICIN BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056, 019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-24

```

### Best Local

Query Match	82.8%;	Score 2101.5;	DB 11;	Length 428;
Best Local Similarity	99.8%;	Pred. No. 3.5e-99;		
Matches 427;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;

Qy	38	ENGAIAQVPPSSNRANESQAEQOEOPKULDSEPRDKARKEVEEYVKTI VGESEYASTKRRH	97
Ds	1	ENGAIAQVPTSSNRANESQAEQOEOPKULDSEPRDKARKEVEEYVKTI VGESEYASTKRRH	60
Qy	98	TIITVALVNEINNINKIEYLINKI VESTSESOLQIUMMESRSKDEAVSKREKDSSSSSSDS	157
Ds	61	TIITVALVNEINNINKIEYLINKI VESTSESOLQIUMMESRSKDEAVSKREKDSSSSSSDS	120
Qy	158	STPRPASDTAKPRPKPTPEPEKVAEAKKVEEA- KKAHQKEEDRRNPTIITYKTLEIEIA	216
Ds	121	STPRPASDTAKPRPKPTPEPEKVAEAKKVEEAEKKAKQKEEDRRNPTIITYKTLEIEIA	180
Qy	217	ESDVEYKKAELIELYVKKANEPREDEQIKQAEAEVESKQAEATRLKIKITDREAEAEBAKR	2767
Ds	181	ESDVEYKKAELIELYVKKANEPREDEQIKQAEAEVESKQAEATRLKIKITDREAEAEBAKR	240
Qy	277	RADAKEOGPRKGARCVGPELATPDKKENDAKSSDSVGEETLPSPSLKPREKKVAEAEK	336
Ds	241	RADAKEOGPRKGARCVGPELATPDKKENDAKSSDSVGEETLPSPSLKPREKKVAEAEK	300
Qy	337	KVEAEAKKKAEDQKEEDRRNPTNTYKTLEIEIAESDVEYKKAELIELYKSEAKEPRENEKY	396
Ds	301	KVEAEAKKKAEDQKEEDRRNPTNTYKTLEIEIAESDVEYKKAELIELYKSEAKEPRENEKY	360
Qy	397	KQAKAEVESKKAEBATRLKIKITDRKKAEEAKRKAEEYKVEKPAEQOPAPAPKAEXP	456
Ds	361	KQAKAEVESKKAEBATRLKIKITDRKKAEEAKRKAEEYKVEKPAEQOPAPAPKAEXP	420
Qy	457	APAPKPEN 464	
Ds	421	APAPKPEN 428	

## RESULT 11

US-09-056-019-1  
Sequence 1, Application US/09056019A  
Publication No. US20030096950A1  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzemann, Theresa  
APPLICANT: Maasure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL

```

; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFOR AND USES THEREOF
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-1

Query Match          78.0%; Score 1979.5; DB 11; Length 406;
Best Local Similarity 99.8%; Pred. No. 4.9e-93;
Matches 405; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 38 ENEGATQVPTSSNRANESQAEQGPCKLDSERDARKAVEEYVKKIYGESYAKSTKRRH 97
Db 1 ENEGATQVPTSSNRANESQAEQGPCKLDSERDARKAVEEYVKKIYGESYAKSTKRRH 60

Oy 98 TITVALVNLNNIKNEYINKIYESTSESQLOIIMESRSKYDEAVSKTEKSSSSSSDS 157
Db 61 TITVALVNLNNIKNEYINKIYESTSESQLOIIMESRSKYDEAVSKTEKSSSSSSDS 120

Oy 158 STKEPASTPAKPNKTEPGEKVAEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLELIA 216
Db 121 STKEPASTPAKPNKTEPGEKVAEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLELIA 180

Oy 217 ESDVEVKKAELVVKANEPREDEQIKQAEAEVSKQAEATRLKKITDREAEAEAKR 276
Db 181 ESDVEVKKAELVVKANEPREDEQIKQAEAEVSKQAEATRLKKITDREAEAEAKR 240

Oy 277 RADAKEQCKPGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKYAEAEK 336
Db 241 RADAKEQCKPGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKYAEAEK 300

Oy 337 KVEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVVKKAELVKEAEKPEPNEKRV 396
Db 301 KVEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVVKKAELVKEAEKPEPNEKRV 360

Oy 397 KQAKAEVSKKAEATRLKIKITDRKKAEEAEKRAKAEEDKYAEKPA 442
Db 361 KQAKAEVSKKAEATRLKIKITDRKKAEEAEKRAKAEEDKYAEKPA 406

RESULT 12
US-09-298-523B-9
; Sequence 9, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-9

Query Match          69.4%; Score 1761.5; DB 11; Length 487;
Best Local Similarity 71.5%; Pred. No. 6.3e-82;
Matches 376; Conservative 44; Mismatches 51; Indels 55; Gaps 8;

Oy 1 MFASKERKHYHSIRKFSVG-ASVVVASLVGSGVVAHTENEGATQVPTSSNRANESQAEQ 59
Db 1 MFASKERKHYHSIRKFSIGVASVVAASLVFGSVVAHTEKEVTQTQVATSPFNANKSQ--- 57
```

```

Oy 60 GEOPCKLDSERDARKAVEEYVKKIYGESYAKSTKSHITITVALVNLNNIKNEYINKIV 119
Db 58 -----TEHMKAAKQVDEIYTKKL-----QDRRHGTQVGLTKGVIKTEYLRHS 104

Oy 120 ESTSESQLOIIMESRSKYDEAVSKFEKSSSSSSDSSSTKPEASDTAKPNKPTPEGKRV 179
Db 105 VSKKESAE-LPSEIKAKLDAFQPKD-----LPTPEGKRV 142

Oy 180 AEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLEIAESDVVKKAELVVKANEP 238
Db 143 AEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLEIAESDVVKKAELVVKANEP 202

Oy 239 DEQIKQAEAEVSKQAEATRLKKITDREAE-----EAKRRADAKEQ 284
Db 203 MEQVNOAKAVESKQAEATRLKKITDREAEATRLLENITKTDREKAEAKRKADKOD 262

Oy 285 KPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKYAEAEKVEAKKK 344
Db 263 ESKRRVKGVPQATLDPKKENDAKSSDSSVGEETLPSLSKSKYAEAEKVEAKKK 322

Oy 345 AEDQEKEDRRNYPTNYKTLLEIAESDVVKKAELVKEAEKPEPNEKVKQAKAVE 404
Db 323 AKDQEKEDRRNYPTNYKTLLEIAESDVVKKAELVKEAEKPEPNEKVKQAKAVE 382

Oy 405 SKKAEATRLKIKITDRKKAEEAEKRAKAEEDKYAEKPAEQPPAPAPAEKPAKPEPN 464
Db 383 SKKAEATRLKIKITDRKKA-EEAKRAKAESEKVEKRAEQPPAPAPAEKPAKPEPN 441

Oy 465 PAEQPKAEKPADQAEEDYARSEENRLTQOQPPTEKPAQST 510
Db 442 PAEQPKAEKPADQAEEDYARSEENRLTQOQPPTEKPAQST 487

RESULT 13
US-09-298-523B-66
; Sequence 66, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 66
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-66

Query Match          68.8%; Score 1745.5; DB 11; Length 487;
Best Local Similarity 71.3%; Pred. No. 4.1e-81;
Matches 376; Conservative 45; Mismatches 49; Indels 57; Gaps 10;

Oy 1 MFASKERKHYHSIRKFSVG-ASVVVASLVGSGVVAHTENEGATQVPTSSNRANESQAEQ 59
Db 1 MFASKERKHYHSIRKFSIGVASVVAASLVFGSVVAHTEKEVTQTQVATSPFNANKSQ--- 57

Oy 60 GEOPCKLDSERDARKAVEEYVKKIYGESYAKSTKSHITITVALVNLNNIKNEYINKIV 119
Db 58 -----TEHMKAAKQVDEIYTKKL-----QDRRHGTQVGLTKGVIKTEYLRHS 104

Oy 120 ESTSESQLOIIMESRSKYDEAVSKFEKSSSSSSDSSSTKPEASDTAKPNKPTPEGKRV 179
Db 105 VSKKESAE-LPSEIKAKLDAFQPKD-----LPTPEGKRV 142

Oy 180 AEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLEIAESDVVKKAELVVKANEP 238
Db 143 AEAKKVEEA-KKAKDKEEDRRNYPTIYKTLLEIAESDVVKKAELVVKANEP 202

Oy 239 DEQIKQAEAEVSKQAEATRLKKITDREAE-----EAKRRADAKE-Q 283
```

Db	203	NEQVKNQAKAYESQOAEATRLKKTITDREQOAEATRLNKTIDREKAEAKKAAAEV	262
Qy	284	GKPKRARAGVGBELATPDCKKENDKSSDSVSGEETLSPSLKPCKYVAEAKKYEAAX	343
Db	263	DKLKRITGAIVGEEPAITPDCKKENDKSSDSVSGEETLSPSLKSGKVAEAKKYAEAK	322
Qy	344	KAEDQKEHDRNNYPNTYKLTLELEIAESDVEYKKELELYEAEKPENEEKVQAKAEV	403
Db	323	KAKQKEDDRNNYPNTYKLTLELE-AESDVKYKEAELELYEAEAESNEEKVQAKAEV	381
Qy	404	ESKKAETRLKIKTDRKKAEEAKRKAAEEDVKAEPAEQOPAPAPAEKPADAPAE	463
Db	382	ESKKAETRLKIKTDRKKA-EBAKKAABEDVKAEKPAEQOPAPAPAEKPADAPAE	440
Qy	464	NPABQPAKEKPADQAAEEDYARSRSEENRLLTQOOPPTKEPAQOST	510
Db	441	NPABQPAKEKPADQAAEEDYARSRSEENRLLTQOOPPTKEPAQOST	487

```

RESULT 14
US-09-298-523B-56
: Sequence 56, Application US/09298523B
: Publication No. US20030059438A1
: GENERAL INFORMATION:
: APPLICANT: BRILES et al.
: TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
: FILE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
: FILE REFERENCE: 454312-3140
: CURRENT APPLICATION NUMBER: US/09/298,523B
: CURRENT FILING DATE: 1999-04-23
: NUMBER OF SEQ ID NOS: 78
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 56
: LENGTH: 581
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
: US-09-298-523B-56

```

Query Match	67.7%	Score 1718:	DB 11:	Length 581:
Best Local Similarity	70.3%	Pred. No. 1.2e-79:		
Matches 369:	Conservative 49:	Mismatches 49:	Indels 58:	Gaps 10:
QY	1	MFASKSEKHYATSRKFSVG-ASVYVSLWGSVYHATENEGATQVPTSNRANESQAEQ	59	
DB	1	MFASKSEKHYATSRKFSIGVASVAVSLWGSVYHATENEGSTQAAISSMAK-----	54	
QY	60	GEQPKLDSERDQARKE-VVEEYVKVIGESYAKSTKKRHITTVALVNLNINKOEYLNK-	117	
DB	55	-----TEHRKKAQGVDEYIEKMLE--TQDRRGHTQNVALNTIKLSATIKTYLREL	104	
QY	118	-IVSTGESQQLIMMESKRVDEAVSKFEEDSSSSSSSDSSTPEASDTAKPKKPTBERG	176	
DB	105	NVLSEKSKDELPP--SETKALDLAAFEKFKDT-----LKPQ	138	
QY	177	EKVAEAKKKVVEA-KKAKDQKEEDRRNPPTTYKTLLEIAESDVEYKKAELVKKAN	235	
DB	139	EKVAEAKKKVVEAKKADQKEEDRRNPPTTYKTLLEIAEPDYKKAELVKEBAK	198	
QY	236	EPDEQKIKQAEAVESKQAEATYLUKKIKTDRBEAEFEBAKRAADK-----EQGKP	286	
DB	199	ESRNEGTIKQKEKEVESKQAEATLENTIKTDRKAAEEBAKKAADAKLEAVATSDQGP	258	
QY	287	KGRKRGVPGGLATPPDKKENAKSSDSVGEETLPSPLKREKKVAAEKKVVEBAK	346	
DB	259	KGRKRGVPGGLATPPDKKENAKSSDSVGEETLPSSSLKCGKVAEKKVVEBAKKA	318	
QY	347	DQKEEDRRNPPTNYKTLLEIAESDVEYKKAELVKEBAKPRNEBKVQKQAEAYESK	406-	
DB	319	DQKEEDRRNPPTNYKTLLEIAESDVYKKAELVLEIYKEBAKEPRDEKTIKQAKAYESK	378	
QY	407	KAEATREKIKTDKKAEEBAKQKAEEDKYKEKPAEQOPAPAPKAEKPAAPKPENPA	466	

Db 379 KAETRLLENITKTRDKAAEEZAKRKAADKDYKAKABEQPPAPATOPK-PAKPEKPA 436

QY 467 EQPKAEKPADQQAEDYARRSSEENRLTQQQPPKTEKAQBSTP 511

Db 437 EQPKAEKTDQQAEDYARRSSEENRLTQQQPPKTEKAQCFTP 481

```

RESULT 15
US-09-298-523B-10
; Sequence 10, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE OF INVENTION: AND STRAINS THEROPF AND USES THEREFOR
; FILE REFERENCE: 454312-1140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-10

```

Query Match	67.6%	Score 1715;	DB 11;	Length 483;
Best Local Similarity	70.3%	Pred. No. 1,4e-79;		
Matches 369;	Conservative 48;	Mismatches 50;	Indels 58;	Gaps 10
Qy	1	MFASKSEKRVHYSIKRSFVSG-ASVVAVSLVMSGVVHATENEGATQVPTSSNFRANESQAOE	59	
Db	1	MFASKSEKRVHYSIKRSFSGVAVASVAVSLVMSGVVHATENEGSTQATSTNNAK-----	54	
Qy	60	GEQCKLIDSEEDKARKE-VVEEYVKKIIVGESYASTKTKGHHTTVALVNLNLTNNKNETLAK- 117		
Db	55	-----TEHRKAKAQVVDVEYIEKMLE--IQDLRHHTQVVALNLIKLSATKTYLREBL 104		
Qy	118	-IVSTESGSLQILMMEGRSKVDENAVSKFEKDSGSSSSSSSSDSTKPEASDTAKPNKPTBEG 176		
Db	105	NVLEEKSKDEPL---SEIKAKLIDAAFEKFKDQ-----LAKG 138		
Qy	177	EKVAEAKKKVVEA-KKAKDQKEEDRRNPPTITYKTLLEIAESDVEYKKAELVELYVKAN 235		
Db	139	EKVAEAKKKVVEAKKKAKADQKEEDRRNPPTITYKTLLEIAEFVDYKVAEAEELVYEAK 198		
Qy	236	EPREDOQIKQAEAVESKQAEATRLKKIKITRDREAEAEERAKRADAK-----EQGR 286		
Db	199	ESRNEGITKQAEKEVESKKAEBATLENKITDRKAAEEAKKKADAKLKEANVAATSDQGR 258		
Qy	287	KGRPKRGVPGELAPPODKKENAKSSDSSVGSETLPSPGLKBEKKVAEAKKYAEAKKAE 346		
Db	259	KGRPKRGVPGELAPPDKKENAKSSDSSVGSETLPSSSLKGGKVAEAKKYAEAKKAEKAK 318		
Qy	347	DQKEEDRRNPPTNYKTLLEIAESDVEYKKAELVELYVEEAKKEPNEEEKVQAKAEVESK 406		
Db	319	DQKEEDRRNPPTNYKTLLEIAESDVKVKAELVELYVEEAKKEPDEEETKQAKAEVESK 378		
Qy	407	KAEATRLLEKIKTDKKAEEBAKRAAEEDKYKEKPAEQPOPAKPAKAEKPAKPAKBPEN 466		
Db	379	KAEATRLLENKITDRKKAEEBAKRAAEEDKYKEKPAEQPOPAKPAKPEK--PAPRPEKA 436		
Qy	467	EQPAEAKPADQQAEEEDYARRSEETYNRLTQQQPKYTEKPAOPSTP 511		
Db	437	EQPAEAKTDDQQAEEEDYARRSEETYNRLTQQQPKYTEKPAOPPTP 481		

Search completed: November 21, 2003, 13:42:42  
Job time : 40.1939 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:33:52 ; Search time 13.8061 Seconds  
(without alignments)  
2380.162 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442

Perfect score: 889

Sequence: 1 IKTDREAEAEKRRADAKE.....AEEAKKKAEDKVEKKA 180

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA:\*  
2: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	889	100.0	406 11 US-09-056-019-1	Sequence 1, Appl1
2	889	100.0	428 11 US-09-056-019-24	Sequence 24, Appl1
3	889	100.0	446 12 US-10-254-995-9	Sequence 9, Appl1
4	889	100.0	453 9 US-09-765-272-38	Sequence 38, Appl1
5	889	100.0	460 11 US-09-056-019-38	Sequence 13, Appl1
6	889	100.0	511 11 US-09-298-523B-13	Sequence 13, Appl1
7	889	100.0	513 11 US-09-298-523B-12	Sequence 12, Appl1
8	889	100.0	655 11 US-09-056-019-2	Sequence 2, Appl1
9	889	100.0	693 12 US-09-769-787-185	Sequence 185, App
10	884	99.4	446 12 US-10-254-995-6	Sequence 6, Appl1
11	884	99.4	459 11 US-09-056-019-39	Sequence 39, Appl1
12	861	96.9	451 11 US-09-298-523B-67	Sequence 67, Appl1
13	803.5	90.4	376 11 US-09-056-019-7	Sequence 7, Appl1
14	803.5	90.4	413 11 US-09-056-019-35	Sequence 35, Appl1
15	803.5	90.4	414 12 US-10-254-995-16	Sequence 16, Appl1

16	803.5	90.4	581 11 US-09-298-523B-56	Sequence 56, Appl1
17	803.5	90.4	631 12 US-09-829-382-25	Sequence 25, Appl1
18	803.5	90.4	663 11 US-09-298-523B-58	Sequence 58, Appl1
19	803.5	90.4	663 11 US-09-056-019-8	Sequence 8, Appl1
20	803.5	90.4	663 12 US-09-969-748C-14	Sequence 14, Appl1
21	803.5	90.4	663 12 US-09-949-079-104	Sequence 104, App
22	800.5	90.0	483 11 US-09-298-523B-10	Sequence 10, Appl1
23	795	89.4	439 11 US-09-056-019-30	Sequence 30, Appl1
24	793	89.2	419 11 US-09-056-019-31	Sequence 31, Appl1
25	793	89.2	444 11 US-09-298-523B-5	Sequence 5, Appl1
26	792	89.1	406 12 US-10-254-995-18	Sequence 18, Appl1
27	792	89.1	425 12 US-10-254-995-11	Sequence 11, Appl1
28	791.5	89.0	414 12 US-10-254-995-10	Sequence 10, Appl1
29	789	88.8	481 11 US-09-298-523B-6	Sequence 6, Appl1
30	784	88.2	428 12 US-10-254-995-7	Sequence 7, Appl1
31	784	88.2	439 11 US-09-056-019-37	Sequence 37, Appl1
32	784	88.2	496 11 US-09-298-523B-4	Sequence 4, Appl1
33	777.5	87.5	437 11 US-09-298-523B-64	Sequence 64, Appl1
34	775.5	87.2	412 12 US-10-254-995-17	Sequence 17, Appl1
35	775.5	87.2	425 11 US-09-056-019-36	Sequence 36, Appl1
36	771.5	86.8	439 11 US-09-056-019-28	Sequence 28, Appl1
37	770.5	86.7	424 12 US-10-254-995-14	Sequence 14, Appl1
38	770.5	86.7	426 12 US-10-254-995-12	Sequence 12, Appl1
39	770.5	86.7	437 11 US-09-056-019-29	Sequence 29, Appl1
40	769.5	86.6	425 12 US-10-254-995-13	Sequence 13, Appl1
41	769.5	86.6	437 11 US-09-056-019-32	Sequence 32, Appl1
42	768	86.4	419 12 US-10-254-995-15	Sequence 15, Appl1
43	768	86.4	433 11 US-09-056-019-33	Sequence 33, Appl1
44	751.5	84.5	427 11 US-09-056-019-34	Sequence 34, Appl1
45	743	83.6	487 11 US-09-298-523B-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-056-019-1  
Sequence 1, Application US/09056019A  
Publication No. US20030096950A1  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzmann, Theresa  
APPLICANT: Maasure, H. R.  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056.019A  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-1

Query Match 100.0%; Score 889; DB 11; Length 406;  
Best Local Similarity 100.0%; Pred. No. 5.1e-53;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IKTDREAEAEKRRADAKEQKPKRAGVGEIATPDKENDAKSSDSVGEETLPS 60  
Db IKTDREAEAEKRRADAKQKPKRAGVGEIATPDKENDAKSSDSVGEETLPS 266  
QY 61 PSKPEKKAABAKKYEAKKAEDQKEDRRNYPNTYKTTLEIAESVVEKKLEL 120  
Db 287 PSKPEKKAABAKKYEAKKAEDQKEDRRNYPNTYKTTLEIAESVVEKKLEL 346  
QY 121 VKKEAPEPRKEEVKAKAEVSKKAEATLEIKTKRKAEEAKRKAEEPKVEKPA 180

Db 347 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 406

RESULT 2

US-09-056-019-24

Sequence 24, Application US/09056019A

Publication No. US20030096950A1

GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I

APPLICANT: Wizemann, Theresa

APPLICANT: Masure, H. R.

APPLICANT: Johnson, Leslie S.

APPLICANT: Koenig, Scott

TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL

TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED

FILE REFERENCE: 1340-1-017

CURRENT APPLICATION NUMBER: US/09/056,019A

CURRENT FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 39

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 24

LENGTH: 428

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-056-019-24

Query Match 100.0%; Score 889; DB 11; Length 428;

Best Local Similarity 100.0%; Pred. No. 5,4e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 60

Qy 227 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 286

Qy 61 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 287 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 346

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 180

Db 347 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 406

RESULT 3

US-10-254-995-9

Sequence 9, Application US/10254995

Publication No. US20030138447A1

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/10/254,995

CURRENT FILING DATE: 2002-09-25

PRIOR APPLICATION NUMBER: US/09/286,981

PRIOR FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 9

LENGTH: 446

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-10-254-995-9

Query Match 100.0%; Score 889; DB 12; Length 446;

Best Local Similarity 100.0%; Pred. No. 5,7e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 60

Db 25 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 84

Qy 61 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 85 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 144

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 180

Db 145 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 204

Db 228 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 287

Qy 61 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 288 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 347

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 180

Db 348 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 407

RESULT 4

US-09-765-272-38

Sequence 38, Application US/09765272

Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 453 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-09-765-272-38

Query Match 100.0%; Score 889; DB 9; Length 453;

Best Local Similarity 100.0%; Pred. No. 5,7e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 60

Db 25 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKENDAKSSDSSVGEETLPS 84

Qy 61 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 85 PSLKPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 144

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 180

Db 145 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAEBAEKRAAEEDVKKEKPA 204

RESULT 5  
US-09-056-019-38  
; Sequence 38, Application US/09056019A  
; Publication No. US20030096950A1  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wazemann, Theresa  
; APPLICANT: Masure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Koenig, Scott  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; FILE REFERENCE: 1340-1-017  
; CURRENT APPLICATION NUMBER: US/09/056, 019A  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019-38

Query Match  
Best Local Similarity 100.0%; Score 889; DB 11; Length 460;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 242 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 301  
Qy 61 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 120  
Db 302 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 361  
Qy 121 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 362 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 421

RESULT 6  
US-09-298-523B-13  
; Sequence 13, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: Briles et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298, 523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-13

Query Match  
Best Local Similarity 100.0%; Score 889; DB 11; Length 511;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 263 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322  
Qy 61 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 120  
Db 323 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 382  
Qy 121 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 121 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180

Db 383 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 442

RESULT 7  
US-09-298-523B-12  
; Sequence 12, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: Briles et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298, 523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-12

Query Match  
Best Local Similarity 100.0%; Score 889; DB 11; Length 513;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 263 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322  
Qy 61 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 120  
Db 323 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 382  
Qy 121 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 383 VKEAKEPRNEEKYQAKAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 442

RESULT 8  
US-09-056-019-2  
; Sequence 2, Application US/09056019A  
; Publication No. US20030096950A1  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wazemann, Theresa  
; APPLICANT: Masure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Koenig, Scott  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; FILE REFERENCE: 1340-1-017  
; CURRENT APPLICATION NUMBER: US/09/056, 019A  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019-2

Query Match  
Best Local Similarity 100.0%; Score 889; DB 11; Length 655;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286  
Qy 61 PSLEPEKKVAEAEKKVEAKKKAEADQKEEDRRNPPTNTYKTLLEIAESDVEVKAELEL 120

Db 287 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 346  
Qy 121 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 180  
Db 347 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 406

RESULT 9  
US-09-769-787-185  
; Sequence 185, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbjo, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PMC/P21129MO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 185  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-185

Query Match 100.0%; Score 889; DB 11; Length 693;  
Best Local Similarity 100.0%; Pred. No. 9.1e-53;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 265 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 324  
Qy 61 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 120  
Db 325 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 384  
Qy 121 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 180  
Db 385 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 444

RESULT 10  
US-10-254-995-6  
; Sequence 6, Application US/10254995  
; Publication No. US20030138447A1  
; GENERAL INFORMATION:  
; APPLICANT: Wizenmann, Theresa M.  
; APPLICANT: Koenig, Scott  
; APPLICANT: Johnson, Leslie S  
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
; FILE REFERENCE: 469201-396  
; CURRENT APPLICATION NUMBER: US/10/254,995  
; CURRENT FILING DATE: 2002-09-25  
; PRIOR APPLICATION NUMBER: US/09/286,981  
; PRIOR FILING DATE: 1999-04-06  
; PRIOR APPLICATION NUMBER: US 60/085,743  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-254-995-6

Query Match 99.4%; Score 884; DB 12; Length 446;  
Best Local Similarity 99.4%; Pred. No. 1.2e-52;  
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 228 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 287  
Qy 61 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 120  
Db 288 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 347  
Qy 121 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 180  
Db 348 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 407

RESULT 11  
US-09-056-019-39  
; Sequence 39, Application US/09056019A  
; Publication No. US20030096950A1  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wizenmann, Theresa  
; APPLICANT: Masure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; FILE REFERENCE: 1340-1-017  
; CURRENT APPLICATION NUMBER: US/09/056,019A  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019-39

Query Match 99.4%; Score 884; DB 11; Length 459;  
Best Local Similarity 99.4%; Pred. No. 1.3e-52;  
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 241 IKTDREAEAEERAKRADAKGKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 300  
Qy 61 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 120  
Db 301 PSKPEKKVAEAEKKVEAEKKAAEDQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 360  
Qy 121 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 180  
Db 361 VKEAKEPRNEEKVKQAAEAVESKKAETRLKIKTDRKKAEEBAKRAAEEDKVKEXPA 420

RESULT 12  
US-09-298-523B-67  
; Sequence 67, Application US/09298523B  
; Publication No. US20030059438A1  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-1140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 67  
; LENGTH: 451

TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-67

Query Match 96.9%; Score 861; DB 11; Length 451;  
Best Local Similarity 95.7%; Pred. No. 4,6e-51;  
Matches 179; Conservative 0; Mismatches 0; Indels 8; Gaps 2;

QY 1 IKTDREAEAEERADAKGKGRAGVPGELATPDKKENDAKSSDSVSGEETLPS 60  
DB 195 IKTDREAEAEERADAKGKGRAGVPGELATPDKKENDAKSSDSVSGEETLPS 254  
QY 61 PSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDVVKAEIEL 120  
DB 255 PSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDVVKAEIEL 313  
QY 121 VKEAEKPRNEKVKQA-----KAESKKAETRLKIKTDRKKAEEAKRRAEED 173  
DB 314 VKEAEKPRNEKVKQAQAEVSKAEVSKAEATRLKIKTDRKKAEEAKRRAEED 373  
QY 174 VKYKEKA 180  
DB 374 VKYKEKA 380

## RESULT 13

US-09-056-019-7  
Sequence 7, Application US/09056019A  
Publication No. US20030096950A1  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzmann, Theresa  
APPLICANT: Maure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
TITLE OF INVENTION: THEREFROM AND USES THEREOF  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056,019A  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 7  
LENGTH: 376  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-7

Query Match 90.4%; Score 803.5; DB 11; Length 376;  
Best Local Similarity 86.8%; Pred. No. 3e-47;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;  
QY 1 IKTDREAEAEERADAK-----EOKKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 188 IKTRKKAEEAEKRRADAKLEAVNATSDQKPKGRAGVPGELATPDKKENDAKSSDS 247  
QY 52 SVGEETLPSPLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 111  
DB 248 SVGEETLPSPLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 307  
QY 112 EVKAELELVKEAEKPRNEKVKQAQAEVSKAEATRLKIKTDRKKAEEAKRRAE 171  
DB 308 VKAELELVKEAEKPRNEKVKQAQAEVSKAEATRLKIKTDRKKAEEAKRRAE 367  
QY 172 EDKYKEKA 180  
DB 368 EDKYKEKA 376

RESULT 14  
US-09-056-019-35  
Sequence 35, Application US/09056019A

Publication No. US20030096950A1  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzmann, Theresa  
APPLICANT: Maure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
TITLE OF INVENTION: THEREFROM AND USES THEREOF  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056,019A  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 35  
LENGTH: 413  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-35

Query Match 90.4%; Score 803.5; DB 11; Length 413;  
Best Local Similarity 86.8%; Pred. No. 3.3e-47;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEERADAK-----EOKKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 188 IKTRKKAEEAEKRRADAKLEAVNATSDQKPKGRAGVPGELATPDKKENDAKSSDS 247  
QY 52 SVGEETLPSPLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 111  
DB 248 SVGEETLPSPLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 307  
QY 112 EVKAELELVKEAEKPRNEKVKQAQAEVSKAEATRLKIKTDRKKAEEAKRRAE 171  
DB 308 VKAELELVKEAEKPRNEKVKQAQAEVSKAEATRLKIKTDRKKAEEAKRRAE 367  
QY 172 EDKYKEKA 180  
DB 368 EDKYKEKA 376

## RESULT 15

US-10-254-995-16  
Sequence 16, Application US/10254995  
Publication No. US20030138447A1  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/10/254,995  
CURRENT FILING DATE: 2002-09-25  
PRIOR APPLICATION NUMBER: US/09/266,981  
PRIOR FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 16  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-10-254-995-16

Query Match 90.4%; Score 803.5; DB 12; Length 414;  
Best Local Similarity 86.8%; Pred. No. 3.3e-47;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;  
QY 1 IKTDREAEAEERADAK-----EOKKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 189 IKTRKKAEEAEKRRADAKLEAVNATSDQKPKGRAGVPGELATPDKKENDAKSSDS 248

QY	52	SVGEETLSPSPSLKPEKVAEAEKKVBEAKKKAEDQKEDRRNYPNTYKTLLEIAESDV	111
Db	249	SVGEETLSPSSSLKSGKVAEAEKKVBEAKKKAEDQKEDRRNYPNTYKTLLEIAESDV	308
QY	112	EVKKALELVKEEAKPRNEEKVOKAEVESKKAATRIEKITDRKKAEEEAARKAAE	171
Db	309	KVKEAELELVKEEAKPRDEEKIKOKAKVESKKAATRIENIKTDKKAEEEAARKAAE	368
QY	172	EDKVKKEKPA	180
Db	369	EDKVKKEKPA	377

Search completed: November 21, 2003, 13:42:43  
Job time : 14.8061 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:23:06 ; Search time 14.7902 Seconds  
(without alignments)  
1624.773 Million cell updates/sec

Title: US-09-298-523C-13  
Perfect score: 2537  
Sequence: 1 MFAKSRKRVHSIRKFSVG.....NRLTQQQPPKTEKDPSTP 511

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363	14.3	1020	1 NPH_HUMAN	P12036 homo sapien
2	351.5	13.9	1164	1 BAG_STRAG	P27951 streptococc
3	328	12.9	1391	1 MSTI_DROHY	O08696 drosophila
4	328	12.9	6632	1 UN89_CAEEL	O01761 caenorhabdi
5	316	12.5	831	1 NPH_RAT	P16884 rattus norv
6	315.5	12.4	705	1 TRDN_RABIT	Q28820 oryctolagus
7	311	12.3	539	1 M24_STRPY	P12379 streptococc
8	310	12.2	1087	1 NPH_MOUSE	P12946 mus musculu
9	300	11.8	771	1 CALD_CHICK	P12957 gallus gall
10	292.5	11.5	857	1 NPH_CHICK	P16053 gallus gall
11	291.5	11.5	700	1 TRDN_CANFA	P82179 canis famli
12	290	11.4	728	1 TRDN_HUMAN	Q13061 homo sapien
13	284	11.2	644	1 NPH_RABIT	P54938 oryctolagus
14	283	11.2	915	1 NPH_HUMAN	P07197 homo sapien
15	282.5	11.1	848	1 NPH_MOUSE	P08553 mus musculu
16	280.5	11.1	810	1 NPH_BOVIN	O77788 bos taurus
17	280	11.0	1220	1 IF2F_HUMAN	O60841 homo sapien
18	279	11.0	667	1 CYLI_BOVIN	P35662 bos taurus
19	278	11.0	488	1 CYLI_BOVIN	Q28092 bos taurus
20	274	10.8	2464	1 MAPB_MOUSE	P14873 mus musculu
21	273.5	10.8	845	1 NPH_RAT	P12839 rattus norv
22	270.5	10.7	598	1 CYLI_HUMAN	P35663 homo sapien
23	268.5	10.6	2459	1 MAPB_RAT	P35662 rattus norv
24	266	10.5	483	1 M6_STRPY	P08088 streptococc
25	265.5	10.5	492	1 M5_STRPY	P02977 streptococc
26	265.5	10.5	793	1 CALD_HUMAN	Q05682 homo sapien
27	260.5	10.3	1359	1 ATRX_CAEEL	O97460 caenorhabdi
28	260	10.2	1403	1 YDF3_SCHRO	Q10475 echinosach
29	259.5	10.2	2468	1 MAPB_HUMAN	P46821 homo sapien
30	258.5	10.2	2116	1 MTS2_DICDI	P08799 dictyosteli
31	257.5	10.1	678	1 GARP_PLA6F	P13816 plasmodium
32	257.5	10.1	877	1 INCE_CHICK	P33352 gallus gall
33	257.5	10.1	4684	1 PLEI_HUMAN	Q15149 homo sapien

## ALIGNMENTS

RESULT 1.  
NPH\_HUMAN  
ID NPH\_HUMAN STANDARD; PRT; 1020 AA.  
AC P12036  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
DE (Neurofilament heavy polypeptide) (NF-H).  
GN NPH OR NPH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.  
RX NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
RA MEDLINE=88328981; PubMed=3138108;  
RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;  
RT "The structure and organization of the human heavy neurofilament  
RT subunit (NF-H) and the gene encoding it.";  
RL EMBO J. 7:1947-1955(1988).  
CC  
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC OBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NPH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL, X15306; CA333366.1; -  
CC EMBL, X15307; CA333366.1; JOINED.  
CC EMBL, X15308; CA333366.1; JOINED.  
CC EMBL, X15309; CA333366.1; JOINED.  
CC PIR, S00979; OPHUH.  
CC Genew, HGNC:7737; NEFH.  
CC MTM, 162230.-  
CC GO, GO:0005883; C:neurofilament; NAS.  
CC GO, GO:0007399; P:neurogenesis; NAS.  
CC Interpro, IPR001664; IF.  
CC Pfam, PF00038; filament; 1.  
CC PROSITE, PS00226; IF; 1.

KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.  
 FT DOMAIN 1 100 HEAD.  
 FT DOMAIN 101 423 ROD.  
 FT DOMAIN 414 1020 TAIL.  
 FT DOMAIN 101 132 COIL 1A.  
 FT DOMAIN 133 145 LINKER 1.  
 FT DOMAIN 146 244 COIL 1B.  
 FT DOMAIN 245 266 LINKER 12.  
 FT DOMAIN 267 288 COIL 2A.  
 FT DOMAIN 289 292 LINKER 2.  
 FT DOMAIN 293 413 COIL 2B.  
 SQ SEQUENCE 1020 AA; 111780 MW; 1177C9D3DCFD14 CRC64;

Query Match 14.3%; Score 363; DB 1; Length 1020;  
 Best Local Similarity 25.9%; Pred. No. 5.5e-07;  
 Matches 152; Conservative 79; Mismatches 239; Indels 118; Gaps 22;

QY 5 KSERVHYHSIRKFSVGASVVV-----ASLVMSVVHATENEGATQVPTSSNRANESQAQ 59  
 DB 439 KSEKI--KVEKSKETIYVEQTEFYVTEVEEKEKEEGKEEGEBAAG 496  
 QY 60 GEQPKLDSERDKA--RKEVEYK-KIVGESYAKSKRHTTVALVNLNNIKNEYLN 116  
 DB 497 GEEETKSPAEBAASPEKEAKSPVKEAKSPAFAKSPKEAKSPAFAVSPKEAKSP-- 553  
 QY 117 KIVESTSBSQOILIMESRSKVDNAVSKPEKSSSSSSSDSTKREASDTAKPKPT-EP 175  
 DB 554 --AKEBAKSPPAKSPKEBAKSPFAVSPKPAKSPAEBAKSPAFAKSPKEAKSPVKEE 611  
 QY 176 GEKVAEAKKV--EAKKAKADQKEEDRRNYPTTYKLELEIAESDVEYKAELELVKVA 234  
 DB 612 AKSPAFAKSPVKEAKSPAFAVSPKPAKSP-----KEBAKSPKAKSP-----KEBA 660  
 QY 235 NEPRDEQIKQAEAVEK-----QAEATRLKIKITD--REKAEBAKRAADAKGKRG 288  
 DB 661 KSPKPAKSPVAEAKSPAFAKSPVKEAKSPAFAKSPVKEAKSPAFAKSPVKEAKSP 720  
 QY 289 RAKGVPEELTPDCKENDAKSSDSSVGEETLPSLSLPEKVAEAKKVEBAKKAADQ 348  
 DB 721 KAKSPVKEAKTPPEAKSPVKEAKSP-----PEKAKSP-PAKTLDVKSPKPAK 773  
 QY 349 KEEDRRNYPTTYKLELEIAESD-----VEYKAELEL--VKEAEPRNEE----- 394  
 DB 774 ARSPADKPEPAKSPVKEAKSPAFAKSPVKEAKSPAFAKSPVKEAKSPAFAKSP 833  
 QY 395 KVKQAKAEVESKKAATRLKIKITDRKKAEEBAKKAABEDVKKE----- 440  
 DB 834 KVEKPPKAAEEKAPATPKTEKDSKK--EEAPKKAAPKVEKEKPAVEKPKESKVE 891  
 QY 441 ----PAEQOPAPAKAEKP-----APAKPENPAQOPKAE 472  
 DB 892 AKKEBAEKKKVPTPEKAPAKVEKEDAKPKPEKTEVAKKEPDDAKAEPSKPAKKA- 950  
 QY 473 KPADOQAEEDYARSEEEYNRLTQOQPKTEKPA-----QSTP 511  
 DB 951 APEKDTYBEKAKKPEK-----EKTEKAKEDDKTISKERSKP 989

RESULT 2  
 BAG\_STRAG STANDARD; PRT; 1164 AA.  
 AC P2751;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE IGA FC receptor precursor (Beta antigen) (B antigen).  
 GN BAG.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_Taxid=1311;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.

RC STRAIN=IA239;  
 RX MEDLINE=91312121; PubMed=1857207;  
 RA Jernstrem P.G., Chhatwal G.S., Timms K.N.;  
 RT "The IGA-binding beta antigen of the c protein complex of Group B  
 RT streptococci: Sequence determination of its gene and detection of two  
 RT binding regions.";  
 RL Mol. Microbiol. 5:843-849(1991).  
 RN [2]  
 RP IDENTIFICATION OF IGA-LIKE DOMAIN.  
 RX MEDLINE=9703265; PubMed=8880921;  
 RA Bateman A., Eddy S.R., Chochia C.;  
 RT "Members of the immunoglobulin superfamily in bacteria.";  
 RL Protein Sci. 5:1939-1942(1996).  
 CC -I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 CC an amide bond (potential).  
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X59771; CAA42442.1; .  
 DR PIR: S15330; PCSOAG.  
 DR InterPro: IPR004829; Gaurface\_antigen.  
 DR InterPro: IPR005877; Gpos\_YsIRK.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR006192; LPXTG.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF05062; RICH; 1.  
 DR Pfam: PF04650; YsIRK\_signal; 1.  
 DR ProDom: PD153432; Gaurface\_antigen; 1.  
 DR SMART: SM00409; IG; 1.  
 DR TIGRFAMs: TIGR01167; LPXTG\_anchor; 1.  
 DR TIGRFAMs: TIGR01168; YsIRK\_signal; 1.  
 DR PROSITE: PS50847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE: PS50835; IG\_LIKE; FALSE\_NEG.  
 KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; signal;  
 KW Immunoglobulin domain.  
 FT SIGNAL 1 37  
 FT CHAIN 38 1135 IGA FC RECEPTOR.  
 FT PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).  
 FT DOMAIN 434 534 IG-LIKE.  
 FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).  
 FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).  
 FT DOMAIN 827 945 PRO-RICH REPEATS.  
 FT SITE 1132 1136 LPXTG SORTING SIGNAL (POTENTIAL).  
 FT MOD\_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
 SQ SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;  
 Query Match 13.9%; Score 351.5; DB 1; Length 1164;  
 Best Local Similarity 28.2%; Pred. No. 1.6e-06;  
 Matches 143; Conservative 82; Mismatches 177; Indels 105; Gaps 21;  
 QY 1 MFASKSRKHYHSIRKFSVG--ASYVVASLVMSVVHATENEGATQVPTSSNRA---NES 55  
 DB 1 MFKSNYERKMYHSIRKFSVGVAASVAVASLFWGSVAHASELVKDSVKTTEVAAPYPSMA 60  
 QY 56 QAEQGEQPKLDSERDKA-----KVEVEYKIVGESYAKSKRHTTVALVNLNN 109  
 DB 61 QTDGNNSSSELETTTKEIPTTDIKAVEVEKTAETSATDQKKEK---QLOQWKNN 117  
 QY 110 IKNEYLNKIVESTESQOILIMESRSKVDNAVSKPEKSSSSSSSDSTKREASDTAKP 169  
 DB 118 LKNDVDTILSHQKN-----EFKTKIDE-----TNDSDALLEE----- 152  
 QY 170 NKPTPEKGV-AEAKKVEBAKKAADQKEEDRRNYPTTYKLELEIAESDVEV--KKA 226  
 DB 153 NQFMETRLRLHKHEEVEKOKKAKQK-----TLKQSDTVADSNIDKELNHQSK 204



QY 227 LELVVKANBPRDEQIKQAEVAVESKQAEATRLKIKITDRBEAEKARRA-----K 281  
 Db 205 VE-----KMAEQKGTINEDKOSMLKKIKIDIRKQA QAKDKEDAEVKRE 248  
 QY 282 EOGKPKGAKGVPELATPPKENDAKSSSVGEETLPPSPSLPEKKVAEAKVEEA 341  
 Db 249 ELGKLFSSTKAGLDQEIQEHVYKETSSEBNTQKVDHVAHSLQNLQKSLBELDKAT--- 305  
 QY 342 KKKAEQOEEDRRNPTTYTLELE--IAESDVEVKA---ELELVKEAEKPRNE-- 394  
 Db 306 ---TNEQATQVKNQFLENAQKLEIQPLIKETNVLTYKAMESELEQVKEKLE--HNSKAN 360  
 QY 395 -----KVQAKAEVSKKAEATRLKIKITDRKAEAEKAKR---AAEDVKKEKPAOP 445  
 Db 361 LEDVAKSEIVREYEGKLNQSKLPEL---KQLEBAHSLKQVVDFFKRTKTSQV 416  
 QY 446 QPAPAPKAEKPAKPPENPAEPKAE 472  
 Db 417 TPKKRVKDLAA-----NENNOQKIE 437  
 RESULT 3  
 MST2 DROHY STANDARD; PRT; 1391 AA.  
 AC 008696;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Axoneme-associated protein msl101(2).  
 GN MSL101(2).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryoeta;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=9504538; PubMed=7957199;  
 RA Neesen J., Padmanabhan S., Buemann H.;  
 RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid  
 RT motif representing the major component of the sperm-tail-specific  
 RT axoneme-associated protein family Dmsl101 form extended  
 RT alpha-helical rods within the extremely elongated spermatozoa of  
 RT Drosophila hydei.";  
 RL Eur. J. Biochem. 225:1089-1095(1994).  
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTES AND EARLY  
 CC SPERMATIDS.  
 CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
 CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT  
 CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM  
 CC REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X73481; CAA51876.1; -;  
 DR PIR; S51364; S51364.  
 DR FlyBase; FBgn0020733; Dhyd\msl101(2).  
 KM Sperm; Repeat; Multigene family; Polymorphism.  
 FT DOMAIN 332 1268 [KR]-K-X-C-X-X-X-A-K-X-X-X-X-X-X-E.  
 FT SEQUENCE 1391 AA; 159000 MW; 1B2A36F30F46878 CRC64;  
 SQ  
 Query Match 12.9%; Score 328; DB 1; Length 1391;  
 Best Local Similarity 27.0%; Pred. No. 1.3e-05;

Matches 140; Conservative 79; Mismatches 220; Indels 80; Gaps 15;  
 QY 38 ENEGATVPTSSNNANE-SQAEQGEOPRKUDSER---DKAEKEVEYKIVGSSYAKST 93  
 Db 363 EKKAKEELAKKKKKADEKKCEAEANKEKKAEEKKCKEKAKEKKEAAKKE 422  
 QY 94 KKRHTITVALVLENNINKNEYLINKIVESTSSQOLIMESRSYVDEAVSFEDSSSSS 153  
 Db 423 KE-----AAERKCEELAKNIKKAEEK--KCKEAAKKEKAEKKCEELAKK 469  
 QY 154 SSDSTKPEASDTAKPNKPTPEGKVAEAKKVEAKKAKQDEDRNRYPTTYTLE- 212  
 Db 470 IKKAEKKKCEETAKKGEVAKKCEELAKKIKKAEIKKKCKLAKKEKTAEEKCEK 529  
 QY 213 -----LEIAESD-----VEYKAELELVKCANBPRDEQIKQA-----PAEV 250  
 Db 530 AAKKKEAAEKKCEKAKAKKREAEKKEKCKSAKKR--EAAKKEKCEKAKKEKAE 587  
 QY 251 ESKQAEATRLKIKITDRBEAEKARRADAKQKPKGRAK-----RGVPELATPDK 303  
 Db 588 KKKCEAAKKEKVAEAKKCEELAKKIKKAEKKCKEAAKKEKAEKKEKCEELAKK 647  
 QY 304 KENDAKSSSVGEETLPPSLKPE-----KVAAEAKVEEA--KKAEQOK 349  
 Db 648 KAAEKKCKKLAKKEKETAEKKEKCEKAKKREAEKKEKCEAAKKEKAEKKECEAA 707  
 QY 350 EEDRRNPTTYTLELEIAESDVEVKAELVKEAEKPRNEKVKQAEVSKKAE 409  
 Db 708 KKEEAAERKKECEELAKKIK-----KAAEKKCKKLAKKKAQKKNKKGKNGKKA- 760  
 QY 410 ATRLKIKITDRKAEAEKARRAEEDVKKEKPAOPAPAPKAEKPAKPPENPAOP 469  
 Db 761 -----LKKKKREELAKKAAKKEKCKEAAKKEKAEKKEKCKEAKKKE--AEKK 811  
 QY 470 KAEKPADQAEEDYARSEE-EYNRLTQOQPPKTEKPAQ 507  
 Db 812 KCEKTAKKRKEAAEKKEKCEKAKAKKREAEKKEKCEKTA 850  
 RESULT 4  
 UN89 CAEBL STANDARD; PRT; 6632 AA.  
 AC 001761; Q17362;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
 GN UNC-89 OR CO9D1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodetinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=86180278; PubMed=8603916;  
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
 RT assembly, encodes a giant modular protein composed of Ig and signal  
 RT transduction domains.";  
 RL J. Cell Biol. 132:835-846(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofibrilment  
 CC lattice assembly begins with positional cues laid down in the  
 CC basement membrane and muscle cell membrane. UNC-89 responds to



```

Db 1485 KKEKSPDATIN1VEVSESE7TIEKTEITMTMTMHSE-----ESTSVKKEKTEPKVDEX 1538
Qy 151 SSSSDSSTKPEPASDTAK---PNKPE7EGEKVAE-----AKKVEAKKAKADOKEDDR 201
Db 1539 PKSP7TKDKKSPKES7ITEIKSPVKKEKSPKEVEKPPASPTKKEKSPKPEKASPTKSENEV 1598
Qy 202 NYPTTYYTLELE1AESDVEVKAELELVKYKANEPRDQOKIQKQAEVSEKQAE1RPLK 261
Db 1599 KSP7TKEKSPKESVVE-----ELKSPKEKSPKAD-D-KPSPTTKEKSPKESATSDV 1649
Qy 262 KIK7DREAE--EAKRADAKQEQP-----KGRKRGVPGELATPDKKEN 306
Db 1650 KSP7TKEKSPKVEKPTSP7TKKSSPP7TKTDDVEKSP7TKEKSPQ7VEEKPPASPTKE- 1708
Qy 307 DAKSDSDSVGE7TLPSPLKPEKKVALEAKKVEEAKKKAEDOKEDDRNNTNTNYKTIEL 366
Db 1709 --KSPKEKSVVEEVKSPKESPEK--AEKPKSP7TKKEKSPKESAABEVKSP7TKEKSPK 1764
Qy 367 E1ASDVEVKKAELELVK--EAEKPE7NEKVKYQAKAEVSEKKAELRLEK1TDRKA 423
Db 1765 SAEKPKSP7TKESSPVKADDEVS7PKPKK--SPKVEEKPKSP7TKK--PTPEKSA 1819
Qy 424 EEEAK-----RKAAEDPKV--EKPAEQOPAPAPKAEKAPAKPENPAEQ 468
Db 1820 ABE7KSP7TKEKSPSSPTK7GTGDEKESPEKPEKPK-SE7PK-KSPGSGPK-KKUSKS 1876
Qy 469 PKAEKPADQQAED 482
Db 1877 PEAKPKPAPKULTRD 1890

```

RESULT 5	NFH_RAT	ID	NFH_RAT	STANDARD;	PRT;	831 AA.
AC	PI6884;	063368;				
DT	01-AUG-1990	(Rel. 15, Created)				
DT	01-FEB-1994	(Rel. 28; Last sequence update)				
DT	16-OCT-2001	(Rel. 40; Last annotation update)				
DE	Neurofilament triplet H protein (200 kDa neurofilament protein)					
DE	(Neurofilament heavy polypeptide) (NF-H) (fragment).					
GN	NFPH OR NFH.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus					
OX	NCBI_TaxID=10116;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Brain.					
RX	MEDLINE=89065087; Pubmed=3143606;					
RA	Breen K.C., Robinson P.A., Wion D., Anderson B.H.;					
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).					
RT	Identification of putative phosphorylation sites.";					
RL	FEBs Lett. 241:213-218(1988).					
RN	[2]					
RP	SEQUENCE OF 37-831 FROM N.A.					
RX	MEDLINE=88309090; Pubmed=2457365;					
RA	Dactysny A., Pham-Dinh D., Koussei C., Felix J.M., Nussbaum J.L.,					
RA	Jolles P.;					
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and					
RT	in situ detection.";					
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).					
RN	[3]					
RP	SEQUENCE OF 1-89 AND 243-313 FROM N.A.					
RX	MEDLINE=87080760; Pubmed=2878828;					
RA	Robinson P.A., Wion D., Anderson B.H.;					
RT	"Isolation of a cDNA for the rat heavy neurofilament polypeptide					
RT	(NF-H)."					
RL	FEBs Lett. 209:203-205(1986).					
RL	[4]					
RP	SEQUENCE OF 318-831 FROM N.A.					
RX	MEDLINE=89184647; Pubmed=2928342;					
RA	Liedbergung I., Spinner N., Snyder S., Anderson J., Goldgaber D.,					
RA	Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;					

```

RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H) : developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MYOTUBE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -I- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -I- PM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY
CC -I- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
-----
DR EMBL; M37227; AAAA4693.1; ALT_FRAME.
DR EMBL; X13804; CAA32038.1; ALT_FRAME.
DR EMBL; M21964; AAA41695.1; -.
DR EMBL; J04517; AAA41692.1; -.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament_1.
DR PROSITE; PS00226; IF_1.
KW Intermediate filament; Coiled coil; Neutone; Phosphorylation; Repeat.
FT NON TER 1
FT DOMAIN 1
FT FT 276 641 51 x 3 AA TANDDEM REPEATS OF K-S-P.
FT CONFLICT 164 164 L->I (IN REF. 2).
FT CONFLICT 185 185 I->S (IN REF. 2).
FT FT 193 193 L->T (IN REF. 2).
FT CONFLICT 199 199 M->T (IN REF. 2).
FT CONFLICT 346 346 K->N (IN REF. 1).
FT CONFLICT 373 373 A->V (IN REF. 1 AND 4).
FT CONFLICT 482 482 G->E (IN REF. 2 AND 4).
FT FT 485 485 P->S (IN REF. 2).
FT CONFLICT 570 571 RK->KE (IN REF. 2 AND 4).
FT CONFLICT 591 591 P->T (IN REF. 2 AND 4).
FT CONFLICT 727 727 A->V (IN REF. 4).
FT FT 757 759 AAP->GST (IN REF. 4).
FT CONFLICT 769 769 T->L (IN REF. 2).
FT CONFLICT 775 775 R->P (IN REF. 2 AND 4).
SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13BF768 CRC64;

Query March 12.5% Score 316; DB 1; Length 831;
Best local similarity 25.3%; Pred.No.2.3e-05;
Matches 150; Conservative 82; Mismatches 213; Indels 148; Gaps 23;

QY 40 EGATOVPTSSNRAN-----ESQAEGEGEPKLDSEDKARKVEE 79
|||:::|||||
DB 191 EGLPKIPSMSTHIKVSKSEKIKVKSSEKTIVIEBTETELQVTELEEDSDKAQDEE 250
|||:::|||||
QY 80 YKKIIVEGSAKSTTKRHTITVALNELNNIKQENLYNKIYESISQQLIMNESRSKV 139
|||:::|||||
DB 251 EEAEEGEEBAATSPAAEAASPEKETKTSVKEE-----AKSPAFAKSP--AEAKSPA- 301
|||:::|||||
QY 140 EAASKPEKDSASSSSSDSSSTR-----DEADDTAKNMKPT-----PGE----- 177
|||:::|||||
DB 302 EAASPAAEVKSPAAKPAEAYKSPAAEVKSPAAEASPAEAKSPAAEVKSPATYKSPGA 361
|||:::|||||
DY 178 -----KVAEAKKVE-----EAKKARDKEEDRRNPVTITYKTLLELI-- 215

```

[illegible]

```

Db      362 EAKSPAEVKSPEAKSPASVKSPEAKSPAEVKSPEAKSPAEVKSPEAKSPA 421
Qy      216 -----ASDVVKAELELVKKNERDQKTKQAEVKSQ-ASATLK 261
Db      422 EVKSPVTVKSPAEKSPAEVKSPEAKSPAEKSPAEKSPAEKSPAEKSPA 481
Qy      262 KIKTRD-EAESEARRADAKGQKPKRAGVGGELATPPKKN-----DAKS 310
Db      482 GAKPPEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPA 541
Qy      311 SD-----SSVGEETLPSPSLK-----PEKVAEAEKV-----BEAKKA 345
Db      542 PEKASPVKEELKPPAEVKSPEKASPKRKEKSEKKTLDVKSPEAKPAKEAKRPA 601
Qy      346 EDQEKEDRRNTPTNTYKTLLEIASDVVKAELELVKKEKPRNEKYKQAEVKS 405
Db      602 DIRSPEQVKSPEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 655
Qy      406 KKAETRLKIKTKDRK-KAESEA-KKAAEDVKVEKPAEQOPAPAKA-EKPAAPKP 462
Db      656 EKTPTPTKEVSKSKDPAKPAQKPKAEKPKLTKPKDPSGAEKKEEAKKAABEE 715
Qy      463 ENPAE-----QPKAEKPADQAEEDYARSEEFYNLTQQQPPKTEKPAQP 508
Db      716 ETPALGVKEAEKPK-EKAEADAKAKEP-SKPSKE-----KKKEEVPAAP 759

RESULT 6
TRDN_RABIT STANDARD; PRT; 705 AA.
AC Q28820; Q28636; Q28637; Q28643;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC STAIN-NEW Zealand white; TISSUE=Skeletal muscle;
RX MEDLINE=93286104; Pubmed=7685347;
RA Kundson C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
RT "Primary structure and topological analysis of a skeletal muscle-
RT specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL J. Biol. Chem. 268:12646-12654(1993).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC TISSUE=Skeletal muscle;
RX MEDLINE=94298946; Pubmed=8026576;
RA Peng M., Fan H., Kitley T.L., Caswell A.H., Schwartz A.;
RT "Structural diversity of triadin in skeletal muscle and evidence of
RT its existence in heart.";
RL FEBS Lett. 346:17-20(1994).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RC TISSUE=Heart muscle;
RX MEDLINE=96132942; Pubmed=8550602;
RA Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT "Biochemical characterization and molecular cloning of cardiac
RT triadin.";
RL J. Biol. Chem. 271:458-465(1996).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX MEDLINE=96066664; Pubmed=7578102;
RA Fan H., Brandt N.R., Caswell A.H.;
RT "Disulfide bonds, N-glycosylation and transmembrane topology of
RT skeletal muscle triadin.";
RL Biochemistry 34:14902-14908(1995).
CC -!- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE

```

```

CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR.
CC -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-
CC linked.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Comment=Additional isoforms seem to exist;
CC Name=Skeletal 1; Synonyms=ST1;
CC IsoId=Q28820-1; Sequence=Displayed;
CC Name=Cardiac 1; Synonyms=CT1;
CC IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
CC Name=Cardiac 2; Synonyms=CT2;
CC IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
CC Name=Cardiac 3; Synonyms=CT3;
CC IsoId=Q28820-4; Sequence=VSP_004466;
CC Name=Skeletal 2; Synonyms=ST2;
CC IsoId=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
CC VSP_004465;
CC Name=Skeletal 3; Synonyms=ST3;
CC IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC -!- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U31540; AAC48496.1; -
CC EMBL; L10065; AAA31488.1; -
CC EMBL; U31555; AAC48497.1; -
CC EMBL; U34201; AAC48498.1; -
CC PIR; A45990; A45990.
CC Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
CC KW Alternative splicing.
CC FT INIT_MET 0
CC FT DOMAIN 1 46
CC FT TRANSMEM 47 67
CC FT DOMAIN 68 705
CC FT CARBOHYD 74 74
CC FT CARBOHYD 624 624
CC FT VARSPLIC 264 285
CC FT VARSPLIC 264 307
CC FT VARSPLIC 286 705
CC FT VARSPLIC 308 705
CC FT VARSPLIC 415 415
CC FT VARSPLIC 416 424
CC FT VARSPLIC 584 584
CC FT VARSPLIC 585 591
CC FT VARSPLIC 645 705
CC FT VARSPLIC 705 79003 MM; 13AFID84475A1361 CRC64;
CC SEQUENCE 705 AA; 12.4k; Score 315.5; DB 1; Length 705;

```

Best Local Similarity 25.0%; Pred. No. 2.1e-05;  
Matches 135; Conservative 83; Mismatches 195; Indels 127; Gaps 21;

QY 38 ENEGATQVPTSSNRNESAQAGPEPKLUDSERDARKEVEYVYKIVGESYASTKGRH 97  
Db 155 ERKLPKTVVHKEKEKEKEKEKEPEKATKHEKLEKEKEPEKTTVTKERKATKEK- 213  
QY 98 TITVALVNLNNIKNEYLNKIVESTESQLOILMMESRSKVDKAVSKEDSSSSSS-- 155  
Db 214 -IEETKKEVGVQKQVAKKQVAKKQVAKKQVAKKQVAKKQVAKKQVAKKQVAKK 272  
QY 156 ----DSSTKPEASDTAKPKRTE-----PGEKVAE--AKKVEAAKXQK 196  
Db 273 IDIFVHGDLKQGSFAPRPPSPTEQASPTPALPPEKKEBKKAKEKVTETKKAKEK 332  
QY 197 EEDRRNYTITVYKLELEIASDVEVKAELELVKVNKNEPRDECKIQAEAVESKQAE 256  
Db 333 EDARKK-----SEKETDIDMKK-----KEGKSPDTPGTGVTQAA- 370  
QY 257 ATRLKIKITDREAEAEAKR-RAADAKQKPKGRKRGVPGELATPDKENDAKSSDSV 315  
Db 371 -----TKDEKEDSKKAKPAEEDPKGKQKKEKHEPKSTYKHEAA----- 415  
QY 316 GEETLPSPSLKPEKKVAEAEKVEAK-KKAEQKEDDRNYPTVYKLELEIAE--- 370  
Db 416 -----PSEKQAKAKIERKEVSAASTKAVPAKKEK-----TTKVEQETRKKEK 462  
QY 371 -----SDVEY-KKAELEL--VKEAKEPRNEKVKQAKAEVSKKAEATRL- 414  
Db 463 KISSVLKDELTKKEKEVVPASLKEKSETKDEKT--SKPEPDIKKEKGEKVKRKP 520  
QY 415 --KIKTRKKAEEAK-RKAAEDKVKKEPAEQOPAPAPAEKPPAPAKENAEOP-- 469  
Db 521 QPQIKKEKPEQDLMKPEKTLHGKPEKVLKQKAVTTEKGVKPKAKAEHKEKPEPS 580  
QY 470 -KAEKPAD-----QAEEDY---ARRSEEEYNTLQQQPPKTEKPAOPS 509  
Db 581 IKTKPKSTSKGMPVEVSGKKIKESKEIKVAPARRSHQLQVNT-----KAEKPARGS 635

RESULT 7  
M24\_STRPY STANDARD; PRT; 539 AA.

AC P12379:  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE M protein, serotype 24 precursor.  
GN EMM24.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vaughn / Serotype M24;  
RX MEDLINE=8811516; PubMed=3276665;  
RA Mow A.R., Beachey E.H., Burdett V.;  
RT "Molecular evolution of streptococcal M protein: cloning and  
RT nucleotide sequence of the type 24 M protein gene and relation to  
RT other genes of Streptococcus pyogenes";  
RL J. Bacteriol. 170:676-684(1988).  
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES  
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF  
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO  
CC PHAGOCYTOSIS.  
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (Potential).  
CC -1- SIMILARITY: TO OTHER M PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M19031; AAA26874.1; -  
CC PIR; A28549; A28549.  
CC InterPro; IPR005877; Gpos YSIRK.  
CC InterPro; IPR001899; Gram\_pos\_anchor.  
CC InterPro; IPR006192; LPXTG.  
CC InterPro; IPR003345; M\_repeat.  
CC Pfam; PF00746; Gram\_pos\_anchor; 1.  
CC Pfam; PF02370; M; 1.  
CC Pfam; PF04650; YSIRK signal; 1.  
CC PRINTS; PR00015; GPOSANCHOR.  
CC TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
CC TIGRFAMs; TIGR01168; YSIRK\_signal; 1.  
CC PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
CC Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;  
KW Antigen; Coiled coil; Signal.  
FT SIGNAL 1 42  
FT CHAIN 43 508  
FT PROPE 509 539  
FT DOMAIN 118 301  
FT REPEAT 118 152  
FT REPEAT 153 187  
FT REPEAT 188 222  
FT REPEAT 223 257  
FT REPEAT 258 292  
FT REPEAT 293 301  
FT DOMAIN 311 405  
FT REPEAT 311 355  
FT REPEAT 311 380  
FT REPEAT 356 380  
FT REPEAT 381 404  
FT DOMAIN 468 504  
FT SITE 505 509  
FT MOD RES 508 508  
SQ SEQUENCE 539 AA; 58804 MW; B03EDF3AC1B6E9C7 CRC64;

Query Match 12.3%; Score 311; DB 1; Length 539;  
Best Local Similarity 27.5%; Pred. No. 2.5e-05;  
Matches 157; Conservative 90; Mismatches 200; Indels 124; Gaps 27;

QY 4 SKSERKHYISIRKFSVG-ASVVVVASLVMSGVATHEGATQVPTSSNRNESAQAGQ 62  
Db 2 TKNNYNHYSRLKLTGTAASVAVALTVLGAGLVNNTN-USAATRS-----Q 48  
QY 63 PKLUDSERDARK-EVEYVYKIVGESYAKSTKGRHTITVALVNLNNIKNEYLNKIVES 121  
Db 49 TDTLEKQERADPFETIENNTLKLKNSDLSFNKALKDNDLTELBSNAK-EKLKNDKS 107  
QY 122 TSESOQLIMMESR-----SKYDAVSK--FEKSSSSSSDSSTKPE-- 162  
Db 108 LSEKASKIQLELFAKADLEKALEGAMNFSTADSAKITLEAEKAAALARRADLEKALEGA 167  
QY 163 ----ASPTAKPNPTERGGEVAAKAKKVEBAKAKOKE-EEDRRNYPT---ITYKLELE 214  
Db 168 MNFSTADSAK-----ITLLEAKAALFAARAELEKALEGAMNFSTADSAKITLEAE 219  
QY 215 IASDVEVKAELELVKANE-----PRDECKIQAEAE--VESKQAEATRL----- 260  
Db 220 --KAAALARRADLE---KALEGAMNFSTADSAKITLEAEKALEKAROLEKALEGAM 273  
QY 261 -----KKIKT-DREAEAEARRADAKQKPKGRKRGVPGEL-----ATP 301  
Db 274 NFASTADSAKITLEAEKAALEAE-KADLEHQSVLVNANRSLRDADASREAKKQLEAEH 332  
QY 302 DKENDAKSSDSVGEETLPSPSLKPEKKVAEAE-KVEBAKKAEDQKEDDRNYPTNT 360  
Db 333 QKLEONKISSEARQSLRDLDASREAKKQLEHOKLEBQONKISASRQSLRDLDASR 392  
QY 361 YKLELEIASDVEVKAELELVKAEAKEPRNEKVKQAKAEVSK-KAEATRL-EKIKT 418

```

Db      393  EAKQVEKALBEANSKLALEKLNLELE--SKULTERKALQKLEAKALEKXU--  448
Oy      419  DRKKAEEBAKRYKAEDYKVEKPAEQOPAPAPAKAEKPAKPEKPNPAEQPAEKP-ADQ  477
Db      449  -AKQAEELAKLRAGK-----ASDSQTPDAKPGNKAVPGKGAPOAGT  489
Oy      478  QAEBYARSRSEBYNRLTQQCPKTEKPAQ  508
Db      490  KPNQKAPMKE-----TKRQLPSTGTETANP  514

RESULT 8
NFH_MOUSE
ID      NFH_MOUSE      STANDARD:      PRT:  1087 AA.
AC      P19246; Q61959;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1992 (Rel. 23, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Neurofilament triplet H protein (200 kDa neurofilament protein)
DE      (Neurofilament heavy polypeptide) (NF-H).
GN      NFH OR NFH.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89121513; PubMed=3220257;
RA      Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosveid F.,
RA      Mushynski W.:
RT      "Sequence and structure of the mouse gene coding for the largest
RT      neurofilament subunit.";
RL      Gene 68:307-314(1988).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89089138; PubMed=3145094;
RA      Sheidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT      "The structure of the largest murine neurofilament protein (NF-H) as
RT      revealed by cDNA and genomic sequences.";
RL      Brain Res. 464:217-231(1988).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Swiss Webster; TISSUE=Brain;
RA      Carden M.J.;
RL      Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC      AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC      NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC      SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC      -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC      PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC      THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC      INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC      OF AXONAL CALIBER.
CC      -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC      OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC      LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC      COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTIONALITY.
CC      -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC      -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC      TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
RR      EMBL; M24496; AAA39813.1; -
RR      EMBL; M23349; AAA39813.1; JOINED.
RR      EMBL; M24494; AAA39813.1; JOINED.

```

[illegible]

OY 464 NPAOPK---AEKPADQQAEDYARRSREE-----VNRLTQQQPKTE 503  
 DB 936 TPERKOSTAKAKEBAGEKKAIVASEETPAKLGKVEAKPKEKETTTEADDTAK 995  
 OY 504 KPAQPS 509  
 DB 996 EPSKPT 1001  
 RESULT 9  
 CALD\_CHICK STANDARD; PRT; 771 AA.  
 AC PI2957; 003698; Q90756; Q90761; Q92018; Q99230;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Caldesmon (CDM).  
 GN CALD1 OR CAD.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).  
 RC TISSUE=Gizzard, and Oviduct;  
 RX MEDLINE=89340480; PubMed=2760048;  
 RA Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;  
 RT "Cloning and expression of a smooth muscle caldesmon.";  
 RN J. Biol. Chem. 264:13873-13879(1989).  
 [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).  
 RC TISSUE=Gizzard;  
 RX MEDLINE=90026426; PubMed=2803315;  
 RA Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;  
 RT "Primary structure and functional expression of h-caldesmon  
 complementary DNA.";  
 RL Biochem. Biophys. Res. Commun. 164:503-511(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).  
 RC TISSUE=Brain;  
 RX MEDLINE=91093148; PubMed=1824698;  
 RA Hayashi K., Fujio Y., Kato I., Sobue K.;  
 RT "Structural and functional relationships between h- and  
 l-caldesmons.";  
 RN J. Biol. Chem. 266:355-361(1991).  
 [4]  
 RP SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).  
 RX MEDLINE=94071934; PubMed=8250919;  
 RA Hartuna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;  
 RT "Common structural and expressional properties of vertebrate  
 caldesmon genes.";  
 RL Biochem. Biophys. Res. Commun. 197:145-153(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).  
 RC TISSUE=Gizzard;  
 RX MEDLINE=92042686; PubMed=1939602;  
 RA Bryan J., Lee R.;  
 RT "Sequence of an avian non-muscle caldesmon.";  
 RN J. Muscle Res. Cell Motil. 12:372-375(1991).  
 [6]  
 RP SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).  
 RC TISSUE=Brain;  
 RX MEDLINE=94271210; PubMed=8002994;  
 RA Yano H., Hayashi K., Hartuna M., Sobue K.;  
 RT "Identification of two distinct promoters in the chicken caldesmon  
 gene.";  
 RN Biochem. Biophys. Res. Commun. 201:618-626(1994).  
 [7]  
 RP SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Gizzard;  
 RX MEDLINE=89273666; PubMed=2730665;

RA Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;  
 RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of  
 the tropomyosin-binding domain in troponin T.";  
 RL Biochem. Biophys. Res. Commun. 161:38-45(1989).  
 RN [8]  
 RP SEQUENCE OF 498-525.  
 RX MEDLINE=88293484; PubMed=3401222;  
 RA Morner D., Audemard E., Derancourt J.;  
 RT "Identification of a 15 kilodalton actin binding region on gizzard  
 caldesmon probed by chemical cross-linking.";  
 RL Biochem. Biophys. Res. Commun. 154:564-571(1988).  
 RN [9]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE=92041815; PubMed=1939059;  
 RA Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;  
 RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of  
 phosphorylation sites.";  
 RL J. Biol. Chem. 266:19971-19975(1991).  
 RN [10]  
 RP PHOSPHORYLATION OF TYR-27 AND TYR-165.  
 RX MEDLINE=20026923; PubMed=10559276;  
 RA Wang Z., Danielson A.U., Mainle N.J., McManus M.J.;  
 RT "Tyrosine phosphorylation of caldesmon is required for binding to the  
 Shc.Grb2 complex.";  
 RL J. Biol. Chem. 274:33807-33813(1999).  
 CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE  
 REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND  
 NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN  
 FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH  
 INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE  
 TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.  
 THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS  
 POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN. MYOSIN, TWO  
 MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN  
 ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPING.  
 CC -1- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON  
 STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS:  
 Event:Alternative splicing; Named isoforms=3;  
 Name=Gizzard h-cad;  
 IsoId=P12957-1; Sequence=Displayed;  
 Name=Brain l-cad;  
 IsoId=P12957-2; Sequence=VSP\_004152, VSP\_004153;  
 Name=Gizzard l-cad;  
 IsoId=P12957-3; Sequence=VSP\_004153;  
 CC -1- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)  
 IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-  
 MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN  
 NON-MUSCLE TISSUES AND CELLS. NOT EXPRESSED IN SKELETAL MUSCLE OR  
 HEART.  
 CC -1- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-  
 BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-  
 BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL  
 HELICAL REGION IN THE MUSCLE FORMS.  
 CC -1- PTM: Phosphorylated in non-muscle cells. Phosphorylation by CDC2  
 during mitosis causes caldesmon to dissociate from microfilaments.  
 Phosphorylation reduces caldesmon binding to Actin, Myosin, and  
 Calmodulin as well as its inhibition of actomyosin ATPase  
 activity. Phosphorylation also occurs in both quiescent and  
 dividing smooth muscle cells with similar effects on the  
 interaction with Actin and Calmodulin and on microfilaments  
 reorganization (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE CALDESMON FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, J04968; AAA49067.1; -







OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----  
EMBL; X17102; CAA34958.1; -.  
DR EMBL; X05558; CAA29073.1; -.  
DR PIR; S15762; S15762.  
DR InterPro; IPR006821; Filament\_head.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR002957; Keratin\_I.  
DR Pfam; PF04732; filament\_1.  
DR Pfam; PF04732; filament\_1.  
DR PRINTS; PR01248; TYPEKERATIN.  
DR PROSITE; PS00226; IF, 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
KM Glycoprotein.  
FT INIT MET 0 0  
FT DOMAIN 1 98 HEAD.  
FT DOMAIN 99 406 ROD.  
FT DOMAIN 407 857 TAIL.  
FT DOMAIN 99 130 COIL 1A.  
FT DOMAIN 131 143 LINKER 1.  
FT DOMAIN 144 242 COIL 1B.  
FT DOMAIN 243 259 LINKER 12.  
FT DOMAIN 260 281 COIL 2A.  
FT DOMAIN 282 285 LINKER 2.  
FT DOMAIN 286 406 COIL 2B.  
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CONFLICT 546 546 G -> R (IN REF. 2).  
SQ SEQUENCE 857 AA; 95704 MW; 442E0FC6AC6477BB CXC64;

Query Match 11.5%; Score 292.5; DB 1; Length 857;  
Best Local Similarity 25.9%; Pred. No. 0.00017;  
Matches 130; Conservative 66; Mismatches 199; Indels 107; Gaps 18;

DY 35 HATENEAGTVPTSNSRANESQAQGE--PQKIDSEDDARKVEEVVKVIQVESTAKS 92  
Db YAKLTTEAEQCKEAIKRSKEBIAYRROLQSKSIIEBSVGTGSLERQLSDIERHNND 349  
DY 93 TKKHHTTVAVLNENNNIKUNELNKIVESTSESOLQIMMSRSKVDEAVSKFEK---- 147  
Db . 350 LTTYQDTIHQLENLRGTMMAHNLREYD-----LNVKKALDIETIANYKLLEGE 402  
DY 148 DSSSSSSSDSGTKPEASDTAKPNKPTEGKVAFA-----KKKVEAKKAQKOKEEDRRNP 204  
Db 403 ETFRSAFSGSITGCIFTH-----RQSVTIASTKIOTKIEPPKL----- 442  
DY 205 TIYTKTELEIAESDVVKKAEI-----ELVKVANERPDOOKIKQAFA---EYESK 253  
Db 443 KVOHKFVEBEIIIEFKVEBKSEMEDALSALAEEMAAKQEEEOEBEKAEEBAVEEAVSE 502  
DY 254 QAEATRLRKIKTDREBAEBAEKRAADAKEQGPKPGRAKRGVPGELATPDKKENDAKSGDS 313  
Db 503 KAAEQAAEEBEKEBEBAEBAEPAKSDAABEGSK-----KEIIEKKE-- 545  
DY 314 SVGEETLPSPSLRKEKKVYAIEAKKYEAQKKAEPQOKEEDRRNPTNYTKTLELIAESDV 373  
Db 546 --GEEA-----EEBEAEAKGAEEAGAAVEYKVPAPASPSPSPPS---PVTEQAK 592  
DY 374 EVKVALELVYEAKPEPNEEKVQQAAYEVSKKA---EATPLEKITDRKKAEEBAEKRA 431  
Db 593 AVQQAALAEVGCDQAEAKAEAKAEKAAEKAASPEKATKTSTPEKAPLPPEKRPPIE---KA 649

Oy		432 AEDDKVK--EKPAEPOGAPAKPEKAPARPKPENPADQPAKEKDADQAEDDYARRSEE	489
Dd		650 ITPEKVRSPEKPPT-TPEKVVSP- EKVPASPEXRTPEKVASPEKA-----	6922
Oy		490 EYNRLTOOQPTEKEXAOPSTR 511   :        ----TPEKPRTPPEKATPEKR 709	
RESULT 11			
ID TRDN CANFA	STANDARD;	PRT; 700 AA.	
AC P82179:			
DT 28-FEB-2003	(Rel. 41,	Created)	
DT 28-FEB-2003	(Rel. 41,	Last sequence update)	
DT 28-FEB-2003	(Rel. 41,	Last annotation update)	
DE Triadin.			
GN TRDN.			
OC Canis familiaris (Dog).			
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
SC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX NCBI_Taxid=9615;	[1]		
RN RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.			
RC TISSUE=Heart, and Skeletal muscle;			
RX MEDLINE=99428545; PubMed=10497235;			
RA Kobayashi Y.M.; Jones L.R.;	"Identification of triadin I as the predominant triadin isoform expressed in mammalian myocardium." ;		
RT J. Biol Chem. 274:28660-28668(1999)	-I- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY) .		
CC CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic reticulum.			
CC CC -I- ALTERNATIVE PRODUCTS:			
CC Event=Alternative splicing; Named isoforms=3;			
CC Name=Skeletal;			
CC IsoId=P82179-1; Sequence=Displayed;			
CC Cardiac 1;			
CC IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;			
CC Cardiac 3;			
CC Name=Cardiac 3;			
CC IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;			
CC -I- TISSUE SPECIFICITY: SKLETAL AND CARDIAC MUSCLE.			
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb.sib.ch) .			
DR EMBL; AF165916; AAF00222.1; -			
DR EMBL; AF165915; AAF00221.1; -			
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;			
KW Alternative splicing.			
FT INIT MET	0		
FT DOMAIN	1 .. 46		
FT TRANSMEM	47 .. 67		
FT DOMAIN	68 .. 700		
FT CARBOHYD	74 .. 74		
FT CARBOHYD	616 .. 616		
FT VARSP LIC	257 .. 277		
FT FT			
FT VARSP LIC	278 .. 700		
FT FT			
FT VARSP LIC	466 .. 466		
FT FT			
FT VARSP LIC	524 .. 579		
BEKKVAVKATEKAIETVTKPPAKKAHOEKESPTIKTD			

```

FT  KKPKTSKETPEVTES -> GILVOVAVVLANCLFLVQFOODE
FT  ELNVESKVFMRHVLHSHPTSTSPILVISTTCRT (in
FT  isoform Cardiac 3).
FT  /FTID=VSP_004004.
SQ  SEQUENCE 700 AA; 78152 MW; F033E3AA1BEE0C56 CRC64;

Query Match 11.5%; Score 291.5; DB 1; Length 700;
Best Local Similarity 25.2%; Pred. No. 0.00016;
Matches 138; Conservative 86; Mismatches 201; Indels 123; Gaps 26;

OY 37 TENEGATVPTSSNANESQAEOGQP-----KLLDSERDKARKEVEYKIVGESYA 90
DB 128 TDKGIEIEPLPKOKEIHKEKAEKKEPRKILAKVAHKEKEKV-KEKESKKKTHKEKI 186
OY 91 KSTKKRHITVA-----LVNELNINKVELINKIVSTSSQ-LQILMMSRSKV 138
DB 187 EKKEPEPTKVAKEERKAKTEBKIKEVKGQKOEKVPKTAQKVEVQTPPKAKEKEKE 246
OY 139 DEAVSKFE-KDS-----SSSSSSDSTPKPASPDTAKPNKPTPEGEKVAEAKK 184
DB 247 TAAVAKHGOQOYAFRCRYMIDMFVHGDLPPOSPLPPLPTVQASRPT-PASPTLEGE 305
OY 185 KVEEAKKAKDOKEEDRRNYPTTYTTLLEIAESDVEYKKAELLVKYKANEPRDEQIK 244
DB 306 --EEKKKAKEKK-----VTSKTKKE--KEDVKKKSDKOTADIVEKKEPKAPETK 352
OY 245 QAEEAVESKQAEATLKIKITDREAEAEAKRRADAKGQKPKGAKGVGELATPDKK 304
DB 353 QGITKVVQAQA-----AKDKEDSKTKTTPVEE-HKGGKQ-----EKK 392
OY 305 ENDAKSSDSSVGEETLPSPSLKPEKKV-AEAEKVEE-----AKKAEQDOKEEDRRNYPTN 359
DB 393 EKVYEPASSSKKEHAPS-----EKQVAKTEBRAEFTSASTKGAVQKKEEK----- 441
OY 360 TYKTLLELEI-----AESDVEYKKAELLVKEBAKPRNEKVKQAQAEVESKKAETRL 414
DB 442 TTKTVEQEIIRKEKSGKTSTASKDKEPIIKDE-KMPKADKEVKPKRPOSQVKEKSE-S 499
OY 415 KIKDKRK-----KAE-----EAKRAEEDKVEKEAE--OPQAP 449
DB 500 QVKEAKEPEODIAKPEKTVSHGKPEKVKOVKATEKAIETVYKPAKKAHEHOKESP 559
OY 450 APKAEKPAAPK--PE-----NPAPQKAEKPADQQAEDVYARSEEEYNR--LTQOQPPK 501
DB 560 TIKDKPKPKTSKETPEVTESGKKKIEKSEKESKEKAEKHLKEEKVSTRKESLSQSHNTX 619
OY 502 TEKPAOPS 509
DB 620 AEKPARVS 627

RESULT 12
TRDN_HUMAN STANDARD; PRT; 728 AA.
AC Q13061.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taake N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RA "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RT and its localisation to chromosome 6q22-6q23.";
RL Eur. J. Biochem. 233:258-265 (1995).

```

```

CC -1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE
CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC -1- SUBUNIT: Homooligomer of variable subunit number; disulfide-linked
CC (by similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18985; AAA75315.1; -.
CC PIR; S68191; S68191.
CC DR GeneW; HGNC:12261; TRDN.
CC MIM; 603283; -.
CC GO; GO:0016021; C:integral to membrane; TAS.
CC DR GO; GO:0006936; P:muscle contraction; TAS.
CC KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
CC FT INIT MET 0
CC FT DOMAIN 1 46
CC FT TRANSMEM 47 67
CC FT POTENTIAL.
CC FT DISULFID 269 269 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 690 690 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 728 AA; 81423 MW; C1C53BBE1B2A0815 CRC64;

Query Match 11.4%; Score 290; DB 1; Length 728;
Best Local Similarity 23.0%; Pred. No. 0.00016;
Matches 126; Conservative 95; Mismatches 208; Indels 118; Gaps 21;

OY 38 ENEGATVPTSSNANESQAEOGQPKLLDSERDKARKEVEYKIVGESYASTKKRH 97
DB 154 ERKIQTVTHKEKKGKREKREKPRKATHKIEKKEPEPTKY-----AEQKKAK 208
OY 98 TITVALVNELNINKVELINKIVESTSE-SQLQILMMSRSKVD--AVSKFEKSSSSS 153
DB 209 TAEKSEKTKKEVKGQKQKQVTAQKVEVQTPSKPEKEDKEKAQVSGHQDQYAF 268
OY 154 SS-----DSSTPEASDTAKPNKPTPEGEKVAEAKKYVEAKKADQKEEDRRNYPTI 206
DB 269 CRYMIDIFVHGDLPPOSPLPPLPTQASRPTPASPALEE-KEGKKAKEK-----V 322
OY 207 TYKTLLEIAESDVEYKKAELLVKYKANEPRDEQIKQAEEVESKQAEATRLIKITD 266
DB 323 TSETKKKE--KEDIKKSSSEKTAIDVEKKEGKASETKQGITVIAQAQA-----AK 371
OY 267 REAEAEAKRRADAKGQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK 326
DB 372 KDEKKDSKTKTKRAVEAQPGKKQ-----EKKEKVEPAKSPKESHSVPS----- 417
OY 327 PEKVY-AEAEKVEE-----AKKAAEDQKEEDRRNYPTNTYTTLEIAE-----S 371
DB 418 -DKQVKAETERAKEIEIAVSSKKAVPGKKEK-----TTKTVQEIIRKEKSGTSSILK 470
OY 372 DVENKKALELV-----KEBAKPRNEEKVKQAQAEVESK-----KAEATRLKIKITDK 421
DB 471 DKPEPIKGEKVPASLSKEPEPTKKDKMSAGKEVAPKPPQLOGKKEKPEPQIKKAK 530
OY 422 KAESE-----AKKAAEEDKVEKEPAEQOPAPAPAKPAPAKPENPAEQP-- 469
DB 531 PAISEKQIHKQDILVREKTVSHGKPREKVLKQKAVATIEKTAQPKTKKAENHERBPS 590
OY 470 -KAEP-----ADQQAEE--DYARSEEEYN-----RLTQOQPPKT 502
DB 591 IKTKPKPTPKTSKETPEVTESGKKKTEISEKSEKAKDKHILREKVSSTRKESLQAHNTYKA 650

```

Qy 503 EKPAQPS 509  
 Db 651 EKPARVS 657

RESULT 13  
 NFM\_RABIT STANDARD; PRT; 644 AA.  
 ID NFM\_RABIT  
 AC P54538;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DT Neurofilament triplet M protein (160 kDa neurofilament protein)  
 DE (Neurofilament medium polypeptide) (NF-M) (Fragment).  
 GN NEF3 OR NEFM OR NFM  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=97055255; PubMed=8899542;  
 RA Vitadello M., Vettore S., Lamar E., Chien K.R., Gorza L.;  
 RT "Neurofilament M mRNA is expressed in conduction system myocytes of  
 the developing and adult rabbit heart."  
 RL J. Mol. Cell. Cardiol. 28:1833-1844(1996).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: Z47378; CA87454.1;  
 DR PIR: S55395; S55395;  
 DR InterPro: IPR001664; IP.  
 DR Pfam: PF00038; Pfam; 1.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
 KW Glycoprotein.  
 FT NON TER 1  
 FT DOMAIN 1 197 ROD.  
 FT DOMAIN 198 643 TAIL.  
 FT DOMAIN <1 33 COIL 1B.  
 FT DOMAIN 34 50 LINKER 12.  
 FT DOMAIN 51 72 COIL 2A.  
 FT DOMAIN 73 76 LINKER 2.  
 FT DOMAIN 77 197 COIL 2B.  
 FT CARBOHYD 217 217  
 SQ SEQUENCE 644 AA; 72450 MW; 030FDAA622889678 CRC64;

Query March 11.2%; Score 284; DB 1; Length 644;  
 Best Local Similarity 25.2%; Pred. No. 0.00027;  
 Matches 124; Conservative 79; Mismatches 167; Indels 122; Gaps 20;

Qy 57 AEQGEQPKLDSSEDKARKEVEEYVKIVGES-----YAKSTKRRHTTV----- 101  
 Db 88 AEQKKAIR-----SAKEELAEYRQLQSKSIELESVAHMKESLEHNVADIERHNHDL 141

Qy 102 -----ALVNEILNINKNEVLNKIVESTSESQILIMESRSKVDAAVSFEK-----D 148  
 Db 142 SSYODTIQQLLENELRGTKWEVARHLREYD-----LNVKVALDIEIAVRYLLSGEE 194  
 Qy 149 SSSSSSSDSTKPPASDPAKPNKPTPEGEKVAEAKKVE--EAKKADQKEEDRRNYPIT 207  
 Db 195 TRFTFGSITGLPYTH-----RQPSVTISSIQTQKVDAPLTK-----VQ 235  
 Qy 208 YKTELEIAESDVEKKALE-----LVKVAKEPPDEOKIKQAEVSKQAE 257  
 Db 236 HKPVEEIIIEETKVDSEMEDALTAIAELIAVSKKEEKEEAEKGEBOEAEVAA 295  
 Qy 258 TRLKKITDDEEAEEKKARADAKGPKRGAPGELATDDKKNDAKSSDSVGE 317  
 Db 296 KSGPVKATTPPIKKEBEGKEBEGEBEER--EDEGKSDQAEKSGSGSTQVE--GE 351  
 Qy 318 ETLPPSLKPEKKVAEAKKVEAKKAEADQKEEDRRNYPNTYKTELEIAESDVEYK 377  
 Db 352 Q-----EEGETAEGEVEEAEKKEKTEKSE-----EVAAKEEPYTE 390  
 Qy 378 AELELVKEAKEPNEKVKQAKAEVSKKAEATRLK-KITDRKKAEEAKKRAE-- 433  
 Db 391 AKVG-KPEKKKSPVPKSPVEEVK-----PKAEATAGKGEQKEEKEEKKKAKESPK 444  
 Qy 434 EDKVEKPAQOPAPAPAKKAPAKPNPA-----EQPKAKKPADQQAEE 481  
 Db 445 EEKV-EKKEKPKDVPKKKASPVEKEBAEBAATITPTKVGLEKETGEGKPLQOEK 503  
 Qy 482 DYARR---SEEE 490  
 Db 504 EKAGEBGSSEE 515

RESULT 14  
 NFM\_HUMAN STANDARD; PRT; 915 AA.  
 ID NFM\_HUMAN  
 AC P01797;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)  
 DE (Neurofilament medium polypeptide) (NF-M) (Neurofilament 3).  
 GN NEF3 OR NEFM OR NFM  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87275853; PubMed=3608989;  
 RA Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;  
 RT "The human mid-size neurofilament subunit: a repeated protein  
 RT sequence and the relationship of its gene to the intermediate  
 RT filament gene family."  
 RT EMBO J. 6:1617-1626(1987).  
 RN [2]  
 RP SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.  
 RX MEDLINE=88158120; PubMed=2450354;  
 RA Lee V.M.-Y., Ouyos L. Jr., Carden M.J., Holloesi M., Dietzschold B.,  
 RA Lazzarini R.A.;  
 RT "Identification of the major multiphosphorylation site in mammalian  
 RT neurofilaments."  
 RT Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC - SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: Y00067; CAA68276.1; -  
 DR PIR: A27864; A27864.  
 DR GeneW: HGNC:7734; NEF3.  
 DR MIM: 162250; -  
 DR GO: GO:0005883; C:neurofilament; TAS.  
 DR GO: GO:0005200; F:structural constituent of cytoskeleton; TAS.  
 DR InterPro: IPR006821; Filament\_head.  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR002957; Keratin\_1.  
 DR Pfam: PF00038; Filament; 1.  
 DR Pfam: PF04732; Filament; 1.  
 DR PRINTS: PR01248; TYPE1KERATIN.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
 KW Glycoprotein.  
 FT INIT\_MET 0  
 FT DOMAIN 1 103 HEAD.  
 FT DOMAIN 104 411 ROD.  
 FT DOMAIN 412 915 TAIL.  
 FT DOMAIN 104 135 COIL\_1A.  
 FT DOMAIN 136 148 LINKER\_1.  
 FT DOMAIN 149 247 COIL\_1B.  
 FT DOMAIN 248 264 LINKER\_12.  
 FT DOMAIN 265 286 COIL\_2A.  
 FT DOMAIN 287 290 LINKER\_2.  
 FT DOMAIN 291 411 COIL\_2B.  
 FT DOMAIN 613 690 6 X 13 AA TANDDEM REPEATS.  
 FT CAROHD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).  
 FT CAROHD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).  
 SQ SEQUENCE 915 AA; 102317 MW; 5P5D3DF34C9D9E50 CRC64;

Query Match 11.2%; Score 283; DB 1; Length 915;  
 Best Local Similarity 24.0%; Pred. No. 0.0004;  
 Matches 118; Conservative 85; Mismatches 199; Indels 90; Gaps 16;

QY NESQAQGEQPKLDSEDRKARK-----EVEEYVKTIAGESYAKST-----KKR 96  
 DB 367 NENGTQKEMARHREYQDLNVKMAIDIEIAAYRKLLEGERTSTFAGSITGPLYTHR 426  
 QY HTITVALVNELN-----NIKIEYLNKIVESTSQQILMERSKVDAVSKFEEDS 149  
 DB 427 PPIITSSIKIOKTVKAPKLVQHKFVEBIIIEITKVEDSKSEMEALITITELASMKEE 486  
 QY SSSSSSDSTPEAD--TAK--PNKPTPEKVAEAKKVEEKKAKKODKEEDRRYP 204  
 DB 487 KKEAEKEEKEEPEAEVAAKSKVYKATAPVKEEKEEKEEKEEKEEKEEKEEKEE 544  
 QY TITVKTLELAIESDVEVKALELVKVKANEPDEQIKQAEAVESKQAEATLKIK 264  
 DB 545 -----DQAEEGSGSEKESSE-----KEEAGEGEGTEAABGEAEAEKKEKVE 588  
 QY TDREBAEAEARADAKQKPKGAKGVPELTPDKENDAKSSSSVGEETLPSPS 324  
 DB 589 EKSEEVATKEELVADAKVE-KPE-KAKSPVPSPVEEKGKSPVPKSPVEEKGKSPVPKSP 646  
 QY LKPEKVAEAKKVEE-AKKAEDQKEDRRNYPNTYKTLLELAIESDVEVKALELV 383  
 DB 647 VEEKGKSPVPSPVEEKGKSPVSKAKSP-----VPKSPVEAKSKAEVG 696  
 QY KEEAKEPNEEKVKQAKAEVSKKAEATRLKIK--TDKKAEEBAKAKAA----- 432

DB 697 KGEQKE-EKEKVEKAPKEEYKKE-----EKPRDVEKKKAESPVEAEVAEVTITKS 751  
 QY 433 -----EEDTKKEKPADQPPAPAPKAEKPAKPKPENPAPQPAKPADQAEEDVARR 486  
 DB 752 VKVHLKKTKEEGKPLQO-----EKKEKAGGSGSEESGDKAGKSKREDIAVN 802  
 QY 487 SEEEYNRLTQOQ 498  
 DB 803 GEVEKKEVEQS 814

RESULT 15  
 NFM\_MOUSE  
 ID NFM\_MOUSE STANDARD; PRT: 848 AA.  
 AC P08553; Q61961; -  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)  
 DE (Neurofilament medium polypeptide) (NF-M).  
 GN NEF3 OR NEFM OR NFM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8724694; PubMed=3036526;  
 RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;  
 RT "Structure and evolutionary origin of the gene encoding mouse NF-M,  
 RL the middle-molecular-mass neurofilament protein.";  
 RN Eur. J. Biochem. 166:71-77(1987).  
 [2]  
 RP SEQUENCE OF 322-540 FROM N.A.  
 RX MEDLINE=8715637; PubMed=3103856;  
 RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosved F.;  
 RT "Cloning and developmental expression of the murine neurofilament  
 RL gene family.";  
 RN Brain Res. 387:243-250(1986).  
 CC - FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC - FUNCTION: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC - FUNCTION: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC - SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X05640; CAA29127.1; -  
 DR EMBL: M20481; AAA39815.1; -  
 DR PIR: B43772; B43772.  
 DR PIR: B43772; B43772.  
 DR MGI: MGI:97314; Ne13.  
 DR InterPro: IPR006821; Filament\_head.  
 DR InterPro: IPR001664; IF.  
 DR Pfam: PF00038; Filament; 1.  
 DR Pfam: PF04732; Filament; 1.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
 KW Glycoprotein.  
 FT INIT\_MET 0 0

```

FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 432 432 S -> F (IN REF. 2).
FT CONFLICT 539 540 QA -> RR (IN REF. 2).
SQ SEQUENCE 848 AA; 95910 MM; 5F251F274D0F13B6 CRC64;

```

```

Query Match 11.1%; Score 282.5; DB 1; Length 848;
Best Local Similarity 24.9%; Pred. No. 0.00039;
Matches 133; Conservative 79; Mismatches 216; Indels 107; Gaps 21;

```

```

QY 35 HATNEGATVPPTSNNRANESQAGE--QPKLDSERDKARKEVEEYVKIVGSHYAKS 92
DB 294 YAKLTEAEQNKAEIRSAKEIEAYRQLOSKSIELESVGTGKESLERQLSDIERHND 353
QY 93 TKRHTITVALVNEIANKIYESTSESQOILMMESRSKVDAAVSKFEK----- 147
DB 354 LSSVODTIQOLENELRGTKWEMARLREYQD-----LINVKALDIEIAVRLKEGE 406
QY 148 DSSSSSSSSDST-----KPEASPTAKPNKPTBEGEVAAKKEVEAKKADQKEEDR 200
DB 407 ETRFSTFGSITGPLYTHRQPSVTISISKIQTKEVAPKLKVQHKEVEI--IEETKVEDE 464
QY 201 RNYPTITKTELETAESDVEVKKAEELVAVKANEPREDQK-IKQAEAVESKQAEKTR 259
DB 465 KSEMEETITATAEBELAAS---AKEEKEAEKEEPEAKSPVKSPEAKEEEEGE--- 517
QY 260 LKKITDREAEAEBAKR-----RADAKEQKPKGRAGVPGELATPDKENDAKSSDS 314
DB 518 ---KEEBEGQHEEBEGDEGVKSQAEQ---GSEKEG--SEKDEGQHEEBEGTEAEG 569
QY 315 VGEETLPSPSLKPEKVAEAKKVEAKKADQKEEDRNYPTNYTKTLETAESDVE 374
DB 570 EGEEA---EAKKEEKI---EGKVEEVAVKEEIKVEKP-----EKAKSPMP 608
QY 375 VKKAELEIVKEAK---EPNNEKVQAKAIVESKKAETRLKIK-----TDRKKA 424
DB 609 KSPVEEVKPKPEAKAGKEQKEEKEVEEKEKVTKESPKEKEVEKEKPPROVADKKAE 668
QY 425 EBAKKAEE-----DKVK---EKPAEQOPAPAPAKAPAPAPENPAPQPAKEKPAD 476
DB 669 SPVKEKAVEEYITISKVLEKDTKEKQPOKEKVEKA---EEBGGSEEBGSDRSPO 725
QY 477 QQAEDYARNSE---EEYNRLTQQQ-----PPKTEKPAQPS 510
DB 726 ESKKEDIANGVEGKEEEOQTEKSGSGREEKGVLTNGLDVSPAEKKEGSDS 780

```

Search completed: November 21, 2003, 13:32:32  
 Job time : 17.7902 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:22:21 ; Search time 55.4631 Seconds  
(without alignments)  
1462.401 Million cell updates/sec

Title: US-09-298-523C-13  
Perfect score: 2537  
Sequence: 1 MFAKSRKHYHYSIRKFSVG.....NRLTQQQPKTKPKAPQSTP 511

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.GeneSeq\_19Jun03.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2516	99.2	693	23	AAU76890
2	2516	99.2	693	24	ABU02720
3	2516	99.2	694	21	AAV81653
4	2350.5	92.6	655	20	AAV49236
5	2350.5	92.6	655	20	AAV12039
6	2252.5	88.8	460	20	AAV49250
7	2249.5	88.7	460	20	AAV32189
8	2235.5	88.1	459	20	AAV49251
9	2235.5	88.1	459	20	AAV32190

10	2194.5	86.5	446	20	AAV49143	Amino acid sequenc
11	2177.5 <th>85.8</th> <th>446</th> <th>20</th> <th>AAV49140</th> <th>Amino acid sequenc</th>	85.8	446	20	AAV49140	Amino acid sequenc
12	2101.5 <th>82.8</th> <th>428</th> <th>20</th> <th>AAV49238</th> <th>Choline binding pr</th>	82.8	428	20	AAV49238	Choline binding pr
13	2101.5 <th>82.8</th> <th>428</th> <th>20</th> <th>AAV32110</th> <th>Choline binding pr</th>	82.8	428	20	AAV32110	Choline binding pr
14	1979.5 <th>78.0</th> <th>406</th> <th>20</th> <th>AAV49225</th> <th>Polypeptide R2 of</th>	78.0	406	20	AAV49225	Polypeptide R2 of
15	1979.5 <th>78.0</th> <th>406</th> <th>20</th> <th>AAV32098</th> <th>Choline binding pr</th>	78.0	406	20	AAV32098	Choline binding pr
16	1770.5 <th>69.8</th> <th>487</th> <th>23</th> <th>AAU76889</th> <th>Pneumococcal surfa</th>	69.8	487	23	AAU76889	Pneumococcal surfa
17	1730 <th>68.2</th> <th>564</th> <th>19</th> <th>AAV62654</th> <th>C3 binding protein</th>	68.2	564	19	AAV62654	C3 binding protein
18	1726 <th>68.0</th> <th>701</th> <th>23</th> <th>AAU76888</th> <th>Pneumococcal surfa</th>	68.0	701	23	AAU76888	Pneumococcal surfa
19	1718 <th>67.7</th> <th>581</th> <th>20</th> <th>AAV43394</th> <th>S. pneumoniae SpA</th>	67.7	581	20	AAV43394	S. pneumoniae SpA
20	1609.5 <th>63.4</th> <th>488</th> <th>23</th> <th>AAU76887</th> <th>Pneumococcal surfa</th>	63.4	488	23	AAU76887	Pneumococcal surfa
21	1569.5 <th>61.9</th> <th>663</th> <th>20</th> <th>AAV49232</th> <th>ChpA of serotype R</th>	61.9	663	20	AAV49232	ChpA of serotype R
22	1569.5 <th>61.9</th> <th>663</th> <th>20</th> <th>AAV32104</th> <th>Choline binding pr</th>	61.9	663	20	AAV32104	Choline binding pr
23	1569.5 <th>61.9</th> <th>663</th> <th>20</th> <th>AAV43395</th> <th>S. pneumoniae chol</th>	61.9	663	20	AAV43395	S. pneumoniae chol
24	1569.5 <th>61.9</th> <th>663</th> <th>23</th> <th>AAE29201</th> <th>Streptococcus pneu</th>	61.9	663	23	AAE29201	Streptococcus pneu
25	1569.5 <th>61.9</th> <th>663</th> <th>23</th> <th>ABG60639</th> <th>Partial sequence o</th>	61.9	663	23	ABG60639	Partial sequence o
26	1569.5 <th>61.9</th> <th>663</th> <th>24</th> <th>ABP55319</th> <th>Streptococcus pneu</th>	61.9	663	24	ABP55319	Streptococcus pneu
27	1568.5 <th>61.8</th> <th>419</th> <th>20</th> <th>AAV49243</th> <th>N-terminal region</th>	61.8	419	20	AAV49243	N-terminal region
28	1568.5 <th>61.8</th> <th>419</th> <th>20</th> <th>AAV32182</th> <th>N-terminal choline</th>	61.8	419	20	AAV32182	N-terminal choline
29	1512.5 <th>59.6</th> <th>406</th> <th>20</th> <th>AAV49152</th> <th>Amino acid sequenc</th>	59.6	406	20	AAV49152	Amino acid sequenc
30	1506 <th>59.4</th> <th>631</th> <th>18</th> <th>AAV46444</th> <th>ChpA, a choline bi</th>	59.4	631	18	AAV46444	ChpA, a choline bi
31	1504.5 <th>59.3</th> <th>439</th> <th>20</th> <th>AAV49240</th> <th>N-terminal region</th>	59.3	439	20	AAV49240	N-terminal region
32	1504.5 <th>59.3</th> <th>439</th> <th>20</th> <th>AAV32179</th> <th>N-terminal choline</th>	59.3	439	20	AAV32179	N-terminal choline
33	1497.5 <th>59.0</th> <th>437</th> <th>20</th> <th>AAV49244</th> <th>N-terminal region</th>	59.0	437	20	AAV49244	N-terminal region
34	1497.5 <th>59.0</th> <th>437</th> <th>20</th> <th>AAV32183</th> <th>N-terminal choline</th>	59.0	437	20	AAV32183	N-terminal choline
35	1496.5 <th>59.0</th> <th>437</th> <th>20</th> <th>AAV49241</th> <th>N-terminal region</th>	59.0	437	20	AAV49241	N-terminal region
36	1496.5 <th>59.0</th> <th>437</th> <th>20</th> <th>AAV32180</th> <th>N-terminal choline</th>	59.0	437	20	AAV32180	N-terminal choline
37	1494 <th>58.9</th> <th>439</th> <th>20</th> <th>AAV49242</th> <th>N-terminal region</th>	58.9	439	20	AAV49242	N-terminal region
38	1494 <th>58.9</th> <th>439</th> <th>20</th> <th>AAV32181</th> <th>N-terminal choline</th>	58.9	439	20	AAV32181	N-terminal choline
39	1466.5 <th>57.8</th> <th>439</th> <th>20</th> <th>AAV49249</th> <th>N-terminal region</th>	57.8	439	20	AAV49249	N-terminal region
40	1466.5 <th>57.8</th> <th>439</th> <th>20</th> <th>AAV32188</th> <th>N-terminal choline</th>	57.8	439	20	AAV32188	N-terminal choline
41	1465.5 <th>57.8</th> <th>433</th> <th>20</th> <th>AAV49245</th> <th>N-terminal region</th>	57.8	433	20	AAV49245	N-terminal region
42	1465.5 <th>57.8</th> <th>433</th> <th>20</th> <th>AAV32184</th> <th>N-terminal choline</th>	57.8	433	20	AAV32184	N-terminal choline
43	1455.5 <th>57.1</th> <th>426</th> <th>20</th> <th>AAV49146</th> <th>Amino acid sequenc</th>	57.1	426	20	AAV49146	Amino acid sequenc
44	1447.5 <th>57.1</th> <th>425</th> <th>20</th> <th>AAV49147</th> <th>Amino acid sequenc</th>	57.1	425	20	AAV49147	Amino acid sequenc
45	1446.5 <th>57.0</th> <th>424</th> <th>20</th> <th>AAV49148</th> <th>Amino acid sequenc</th>	57.0	424	20	AAV49148	Amino acid sequenc

#### ALIGNMENTS

RESULT 1  
AAU76890  
ID AAU76890 standard; Protein; 693 AA.  
XX  
AC AAU76890;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pneumococcal surface protein C19 TIGR (PspC19 TIGR).  
XX  
KW Hic; factor H; FH; antibacterial; PspC19 TIGR;  
KM pneumococcal surface protein C19 TIGR.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN W0200208426-A2.  
XX  
PD 31-JAN-2002.  
XX  
PF 19-JUL-2001; 2001WO-EP08409.  
XX  
FR 20-JUL-2000; 2000SE-0002738.  
XX  
PA (HANS-) HANSA MEDICAL AB.  
XX  
PI Björck L, Sjöholm A, Janulczyk R, Pozzi G, Iannelli F;  
XX WPI, 2002-257337/30.  
XX  
DR Polypeptide having ability to bind factor H, and proteins and peptides  
XX PT derived from polypeptides useful as vaccine for infections caused by  
XX PT bacteria and for identifying agents inhibiting binding of factor H to

FT		protein -
XX		
PS		Disclousre; Page 59-61; 63pp; English.
CC		The invention relates to the Streptococcus pneumoniae Hic polypeptide
CC		having the ability to bind factor H (FH). The Hic protein and the
CC		polynucleotide encoding it can be used to make vaccine compositions
CC		capable of generating an immune response to Streptococcus pneumoniae or
CC		binding to an anti-protein Hic antibody. The protein and its homologues
CC		are useful for identifying an agent which inhibits binding of factor H to
CC		Streptococcus pneumoniae comprising incubating any one of the
CC		polypeptides with factor H and a test agent, monitoring binding of FH to
CC		the polypeptide and determining thereby whether the test agent inhibits
CC		binding of factor to the polypeptide. This sequence represents the
CC		pneumococcal surface protein C19 TIGR (PepC19 TIGR), a factor H binding
CC		protein.
XX		
SQ		Sequence 693 AA.
	Query Match	99.2%; Score 2516; DB 23; Length 693;
	Best Local Similarity	99.6%; Pred. No. 3,7e-152;
	Matches 511; Conservative	0; Mismatches 0; Indels 2; Gaps 2
OY	1 MPASSEKRVHYSITKRFSGV-ASVVVASLVMGSSVHAATNEGATQVPTSSNPAESQAEO	59
Dd	1 MFASKSEKRVHSIRKFSVGAVSVVAASLVMGSVHAATENEGATQVPTSSNPANESQAEO	60
OY	60 GEOPKKLDSERPDKARVEVEYVKTIYGESYAKSTKRHTTVALVNLNNIKNEYLNKIIV	119
Dd	61 GEOPKKLDSERPDKARKEVEEYVKTIYGESEYASTKKRHITTALVNLNNIKNEYLNKIIV	120
OY	120 ESTSESOLQILMMESRSKVDEAVSKFEKDSSSSSDSSTKPEASDPAKNKPTEBGEKY	179
Dd	121 ESTSESQQLIIMMESRSKVDEAVSKFEKDSOSSSSSDSSTKPEASDPAKNKPTEBGEKY	180
OY	180 AEAKKRVBEA-KKANDQKEEDRRNPPTTYTKTLELFIAASDYEVKKAELELYKVRANER	238
Dd	181 AEAKKRVBEAEKKAADQKEEDRRNPPTTYTKTLELFIAASDYEVKKAELELYKVRANEBR	240
OY	239 DEOKTKOAEVESKOAEATRLPKKITDDEEAABEBAKRRADAKEOGKPGRAKRGVGBEL	298
Dd	241 DEOKTKOAEVESKOAEATRLPKKITDDEEAABEBAKRRADAKEOGKPGRAKRGVGBEL	300
OY	299 ATPDKENDAKSSDSSVGEETLPSPSLKPEKTVAAEKKEVBEAKKKAEDQKEEDRRNPPT	358
Dd	301 ATPDKENDAKSSDSSVGEETLPSPSLKREKTVAAEKKEVBEAKKKAEDQKEEDRRNPPT	360
OY	359 NTYKTLELFIAASDYEVKKAELELYKEAKEPRNEBKVQOAQAEVSKAEATRLFKIT	418
Dd	361 NTYKTLELFIAASDYEVKKAELELYKEAKEPRNEBKVQOAQAEVSKAEATRLFKIT	420
OY	419 DRKKEEBAKRAKAEEDKYKEPAEPPOAPAPKAEKPPAPKPPENAPAPKAEKPADQQ	478
Dd	421 DRKKEEBAKRAKAEEDKYKEPAEPPOAPAPKAEKPPAPKPPENAPAPKAEKPADQQ	480
OY	479 AEDDYARRSEEYNRLLTOQPPTKEPAPQSTP	511
Dd	481 AEDDYARRSEEYNRLLTOQPPTKEPAPQSTP	513
RESULT 2		
ID	ABU02720	standard; Protein; 693 AA.
XX	ABU02720;	
DT	11-FEB-2003	(first entry)
XX		
DE		S. pneumoniae type 4 strain protein from coding region #2259.
XX		
KW		Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection; antiinflammatory; antibacterial; immunostimulant; auditory; respiratory; gene therapy; vaccine.

XX		Streptococcus pneumoniae type 4 strain.
OS		
XX		WO200277021-A2.
PN		
XX		03-OCT-2002.
PD		
XX		27-MAR-2002; 2002WO-IB02163.
PF		
XX		27-MAR-2001; 2001GB-0007658.
PR		
XX		(CHIR-) CHIRON SPA.
PA		(GENO-) INST GENOMIC RES.
XX		
PI	Maignani V, Tetteijn H, Fraser C;	
DR	WPI; 2003-040579/03.	
XX	N-PsDB; ABX08011.	
PT	New proteins and nucleic acid molecules from Streptococcus pneumoniae,	
XX	useful as medicaments for treating or preventing a disease or infection	
PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media	
PT	or ear infection -	
PS	Claim 1; SEQ ID No 4598; 56pp; English.	
XX		
CC	The invention relates to a protein comprising or having at least 50%	
CC	identity to any of the 2469 amino acid sequences, identified in the	
CC	specification (available on a computer readable format), or its fragment,	
CC	expressed from 2469 of 2489 identified DNA coding regions from the	
CC	Streptococcus pneumoniae type 4 strain genomic sequence appearing as	
CC	ABSS66454. Also included are an antibody which binds one of the	
CC	proteins, treating a patient by administering the protein, DNA or	
CC	antibody (in a composition), a kit comprising first and second primers,	
CC	which are the nucleic acid cited above or fragments between nucleotides	
CC	8-100 of a sequence not defined in the specification, for amplifying a	
CC	target sequence contained within a Streptococcus nucleic acid sequence,	
CC	where the first primer is substantially complementary to the target	
CC	sequence and the second primer is substantially complementary to the	
CC	complement of the target sequence, and where the parts of the primers	
CC	having substantial complementarity define the termini of the target	
CC	sequence to be amplified, assay comprising contacting a test compound	
CC	with the protein, and determining whether the test compound binds to the	
CC	protein and a Streptococcus pneumoniae bacterium, where one or more	
CC	genes encoding the proteins has been rendered inactive. The proteins,	
CC	nucleic acid molecules, antibody and compositions are useful as	
CC	medicaments for treating or preventing a disease or infection due to	
CC	streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,	
CC	sepsis, otitis media or ear infection. They are also useful in developing	
CC	vaccines, diagnostics and antibiotics. The methods are useful for	
CC	identifying immunodominant proteins. The present sequence is one of	
CC	the 2469 proteins expressed by the identified coding regions from the	
CC	genomic sequence.	
CC	Note: The sequence data for this patent did not form part	
CC	of the printed specification, but was obtained in electronic	
CC	format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences.	
CC		
XX		
SQ	Sequence 693 AA:	
Query Match	99.2%; Score 2516; DB 24; Length 693;	
Best Local Similarity	99.6%; Pred. No. 3.7e-152;	
Matches 511; Conservative	0; Mismatches 0; Indels 2; Gaps 2	
OY	1 MFASGSEKHYSTRKSFGV-ASVVVASLVNGSVVAHNEGACOVPTSSNRANESQAEO	59
Db	1 MFASGSEKHYSTRKSFGVAVSVVSLVNGSVVAHNEGACOVPTSSNRANESQAEO	60
OY	60 GEOPKLDSEDDKARKVEEYEVKKIVGVSSYAKSTFKRRTITVALNVLNNINIKEXLANKIV	119
Db	61 GEOPKLDSEDDKARKVEEYEVKKIVGVSSYAKSTFKRRTITVALNVLNNINIKEXLANKIV	120
OY	120 ESTSISOLIIIMMESRSKVDEAVSKFEKDSSSSSSDSSTKPASDTRKPKNPTEPGEKV	179



```

Db      121 ESTSESQQLIMMESRSKVDEAVSKFEKDSSSSSSSDSTKEADTAKPNKPTPEGKV 180
QY      180 AEAKKKVEEA-KKADQKDEDRNNPTITTKTLELEIAESDVEYKKALELVKYNANPR 238
Db      181 AEAKKKVEEAERKADQKEBDRNPTITTKTLELEIAESDVEYKKALELVKYNANPR 240
QY      239 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKEOGKPKGAKGVPEL 298
Db      241 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKEOGKPKGAKGVPEL 300
QY      299 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVEBAKKKADQKEBDRNPT 358
Db      301 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVEBAKKKADQKEBDRNPT 360
QY      359 NTYKTLLELEIAESDVEYKKALELVKEAKEPRNEEKYKQAKAVESKKAATRLKIKT 418
Db      361 NTYKTLLELEIAESDVEYKKALELVKEAKEPRNEEKYKQAKAVESKKAATRLKIKT 420
QY      419 DRKKAEEBAKRAAEEDVKKEKPAEQPPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 478
Db      421 DRKKAEEBAKRAAEEDVKKEKPAEQPPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 480
QY      479 AEDYARSRSEEEYNRLTQQQPPKTEKPAQPSPT 511
Db      481 AEDYARSRSEEEYNRLTQQQPPKTEKPAQPSPT 513

```

RESULT 3  
AAV81653  
ID AAV81653 standard; Protein; 694 AA.

AC AAV81653;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID302.

KX Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW antibacterial; antinflammatory; meningitis; infection; diagnosis;

KM pneumococcal disease.

OS Streptococcus pneumoniae.

XX MO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Gilbert CPG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of

XX PT pneumococcal diseases and for screening agents capable of antagonizing

XX or inhibiting expression of the protein

XX Claim 2; Page 96; 108pp; English.

CC AAV81501 to AAV81679 represent specifically claimed protein sequences  
CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent  
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
CC The sequences have antibacterial and antinflammatory properties.  
CC The protein sequences, and fragments of them, are useful as immunogens  
CC and/or antigens. The nucleotide sequences can be used in vaccines and in  
CC diagnostic assays. The proteins and nucleotides can be useful for the  
CC detection and diagnosis of S. pneumoniae. The protein sequences are also

CC useful for screening an agent capable of antagonizing, inhibiting or  
CC interfering with the function or expression of the proteins in which the  
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
CC and meningitis. AA05591 to AA05614 represent primers used in the  
CC exemplification of the present invention.

XX SQ Sequence 694 AA;

Query Match 99.2%; Score 2516; DB 21; Length 694;  
Best Local Similarity 99.6%; Pred. No. 3,76-152;  
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```

QY      1 MFASKSRKHYHSIRKFSVGC-ASYVVASLVWGSVTHATENGATOVPTSSRNANESQAEQ 59
Db      1 MFASKSRKHYHSIRKFSVGCASYVVASLVWGSVTHATENGATOVPTSSRNANESQAEQ 60
QY      60 GEOPKLLDSEDRKARKEVEEYVKKIIGESYAKSTFKKHITITVALVNEINIKNEYLANKIV 119
Db      61 GEOPKLLDSEDRKARKEVEEYVKKIIGESYAKSTFKKHITITVALVNEINIKNEYLANKIV 120
QY      120 ESTSESQQLIMMESRSKVDEAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGKV 179
Db      121 ESTSESQQLIMMESRSKVDEAVSKFEKDSSSSSSSSTKPEASDTAKPNKPTPEGKV 180
QY      180 AEAKKKVEEA-KKADQKDEDRNNPTITTKTLELEIAESDVEYKKALELVKYNANPR 238
Db      181 AEAKKKVEEAERKADQKEBDRNPTITTKTLELEIAESDVEYKKALELVKYNANPR 240
QY      239 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKEOGKPKGAKGVPEL 298
Db      241 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKEOGKPKGAKGVPEL 300
QY      299 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVEBAKKKADQKEBDRNPT 358
Db      301 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVEBAKKKADQKEBDRNPT 360
QY      359 NTYKTLLELEIAESDVEYKKALELVKEAKEPRNEEKYKQAKAVESKKAATRLKIKT 418
Db      361 NTYKTLLELEIAESDVEYKKALELVKEAKEPRNEEKYKQAKAVESKKAATRLKIKT 420
QY      419 DRKKAEEBAKRAAEEDVKKEKPAEQPPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 478
Db      421 DRKKAEEBAKRAAEEDVKKEKPAEQPPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 480
QY      479 AEDYARSRSEEEYNRLTQQQPPKTEKPAQPSPT 511
Db      481 AEDYARSRSEEEYNRLTQQQPPKTEKPAQPSPT 513

```

RESULT 4  
AAV49226  
ID AAV49226 standard; Protein; 655 AA.

AC AAV49226;

DT 07-FEB-2000 (first entry)

DE CBPA of serotype 4 amino acid sequence.

KX Choline binding protein A; CBPA; truncate; immune response; infection;

KW pneumococcal bacterium; vaccine.

OS Streptococcus pneumoniae.

XX MO9951187-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-US07668.

XX 07-APR-1998; 98US-0056019.

XX 07-APR-1998; 98US-0080878.

PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX  
XX Tuomanen EI, Maure HR;  
PI  
XX WPI, 1999-620161/53.  
DR N-PSDB; AA231402.  
XX  
XX Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections -  
XX  
XX Disclosure; Page 9; 85pp; English.  
PS  
XX  
XX The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC AA19225, AA19227, AA19230, AA19231, AA19233 and AA19238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use.  
XX  
XX Sequence 655 AA;  
SQ

Query Match 92.6%; Score 2350.5; DB 20; Length 655;  
Best Local Similarity 99.8%; Pred. No. 1.2e-141;

Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEGQOPKLDSEBDKARKEVEEYVKTIAGESYAKSTKKRH 97  
DB 1 ENEGATQVPTSSNRANESQAEGQOPKLDSEBDKARKEVEEYVKTIAGESYAKSTKKRH 60  
QY 98 TITVALVNLNNIKNEYLINKIVESTSESOQLIMMESRSKYDEAVSKFEKSSSSSSSDS 157  
DB 61 TITVALVNLNNIKNEYLINKIVESTSESOQLIMMESRSKYDEAVSKFEKSSSSSSSDS 120  
QY 158 STKPEASTAKNKPTPEGEKVAEAKKVEEAKKQKEDDRNYPITTYKTLELEIA 216  
DB 121 STKPEASTAKNKPTPEGEKVAEAKKVEEAKKQKEDDRNYPITTYKTLELEIA 180  
QY 217 ESDVEVKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 276  
DB 181 ESDVEVKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 240  
QY 277 RADAKEOGKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 336  
DB 241 RADAKEOGKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 300  
QY 337 KYEEAKKKAADQKEDDRNYPITTYKTLELEIAESDVEVKKAELVLYKBEAKEPNNEKY 396  
DB 301 KYEEAKKKAADQKEDDRNYPITTYKTLELEIAESDVEVKKAELVLYKBEAKEPNNEKY 360  
QY 397 KOAKAEVESKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 456  
DB 361 KOAKAEVESKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 420  
QY 457 APAPKPNPAEQPKAEKPADQAEEDVYARSEEBYNTLQOQPKTEKPAQSTP 511  
DB 421 APAPKPNPAEQPKAEKPADQAEEDVYARSEEBYNTLQOQPKTEKPAQSTP 475

RESULT 5  
AA192099 standard; Protein; 655 AA.  
ID AA192099 standard; Protein; 655 AA.  
XX

AC AA192099;  
XX  
XX 01-FEB-2000 (first entry)  
DT  
XX Choline binding protein A (Cbpa).  
XX  
XX Choline binding protein; Cbpa; adhesin; immunogen;  
XX vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
XX otitis media; pneumonia.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO951188-A2.  
XX  
XX 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US07669.  
XX  
XX 07-APR-1998; 98US-0056019.  
XX  
XX 07-APR-1998; 98US-0080878.  
XX  
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;  
XX WPI, 1999-633690/54.  
XX  
XX N-PSDB; AA234452.  
XX  
XX New N-terminal choline binding protein A truncate polypeptides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections -  
XX  
XX Disclosure; Page 88-93; 160pp; English.  
PS  
XX  
XX This sequence represents the choline binding protein A (Cbpa)  
CC of Streptococcus pneumoniae serotype type 4. The invention provides  
CC novel N-terminal Cbpa truncated polypeptides (see AA192098-110 and  
CC AA192179-90) and polynucleotides encoding them, host-vector systems,  
CC and antibodies that specifically bind to the polypeptides. The  
CC invention also relates to vaccines including the polypeptides,  
CC which provide protection or elicit protective antibodies to  
CC bacterial infection, specifically pneumococcus, and to antibodies  
CC and antagonists against such polypeptides for use in diagnosis and  
CC passive immunotherapy. The polypeptides and/or polynucleotides are  
CC also useful as competitive inhibitors of bacterial adhesion of  
CC pneumococcus.  
XX  
XX Sequence 655 AA;  
SQ

Query Match 92.6%; Score 2350.5; DB 20; Length 655;  
Best Local Similarity 99.8%; Pred. No. 1.2e-141;

Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEGQOPKLDSEBDKARKEVEEYVKTIAGESYAKSTKKRH 97  
DB 1 ENEGATQVPTSSNRANESQAEGQOPKLDSEBDKARKEVEEYVKTIAGESYAKSTKKRH 60  
QY 98 TITVALVNLNNIKNEYLINKIVESTSESOQLIMMESRSKYDEAVSKFEKSSSSSSSDS 157  
DB 61 TITVALVNLNNIKNEYLINKIVESTSESOQLIMMESRSKYDEAVSKFEKSSSSSSSDS 120  
QY 158 STKPEASTAKNKPTPEGEKVAEAKKVEEAKKQKEDDRNYPITTYKTLELEIA 216  
DB 121 STKPEASTAKNKPTPEGEKVAEAKKVEEAKKQKEDDRNYPITTYKTLELEIA 180  
QY 217 ESDVEVKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 276  
DB 181 ESDVEVKKAELVLYKCANEPDEOKIKQAEAEVESKQAEATRLKKITDREAEAEBAKR 240  
QY 277 RADAKEOGKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 336  
DB 241 RADAKEOGKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 300

QY 337 KVEAKKKAEDEDRNRYPTNTYKTLLEIAESDVEVKAELELVKEAEKPEPNEKV 396  
DB 301 KVEAKKKAEDEDRNRYPTNTYKTLLEIAESDVEVKAELELVKEAEKPEPNEKV 360  
QY 397 KQAKAEVSKKAATRLKIKTDKKAEEBAKKAEDKVEKPAEQOPAPAKAEKP 456  
DB 361 KQAKAEVSKKAATRLKIKTDKKAEEBAKKAEDKVEKPAEQOPAPAKAEKP 420  
QY 457 APAPPEPAPQKAEKPADQAEEDYARSRSEBYNRLTQQOPPTTEKPAOPSTP 511  
DB 421 APAPPEPAPQKAEKPADQAEEDYARSRSEBYNRLTQQOPPTTEKPAOPSTP 475

RESULT 6  
AA49250  
ID AA49250 standard; Protein; 460 AA.  
AC AA49250;  
XX  
XX 07-FEB-2000 (first entry)  
DT  
DE N-terminal region of Cbpa polypeptide Ntype4.  
XX  
XX Choline binding protein A; Cbpa; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9951187-A2.  
XX  
PD 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US07668.  
PF  
XX 07-APR-1998; 98US-0056019.  
PR 07-APR-1998; 98US-0080878.  
XX  
XX (SUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA  
XX  
PI Tuomanen EI, Masure HR;  
XX  
DR WPI; 1999-620161/53.  
XX  
PT Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections  
XX  
PS Disclosure; Fig 2A-B; 85pp; English.  
XX  
XX The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC AA49225, AA49227, AA49230, AA49231, AA49233 and AA49238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use.  
XX  
SQ Sequence 460 AA;

Query Match 88.8%; Score 2252.5; DB 20; Length 460;  
Best Local Similarity 99.6%; Pred. No. 1.4e-135;  
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWGSVVHATENEGATVPTSSNRANESQAEQEGOPKLDSEDPKAREVEYVK 83  
DB 2 IVASLWGSVVHATENEGATVPTSSNRANESQAEQEGOPKLDSEDPKAREVEYVK 61  
QY 84 IVGSYAKSTKKRHTITVALVNELNINKBYLNKIVESTSSQOLIMMESRSKYDEAVS 143  
DB 62 IVGSYAKSTKKRHTITVALVNELNINKBYLNKIVESTSSQOLIMMESRSKYDEAVS 121  
QY 144 KFEKSSSSSSSDSTKREASDTAKPNKPTPEGEVVAEAKKVEBA-KKADQKEDERN 202  
DB 122 KFEKSSSSSSSDSTKREASDTAKPNKPTPEGEVVAEAKKVEBAKKAADQKEDERN 181  
QY YPTITTYKTLLEIAESDVEVKAELELVKVAANEPRDQKIKQAEVSKQAEATRLKK 262  
DB 182 YPTITTYKTLLEIAESDVEVKAELELVKVAANEPRDQKIKQAEVSKQAEATRLKK 241  
QY 263 IKTDREAEBAEKRRADAKQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 322  
DB 242 IKTDREAEBAEKRRADAKQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 301  
QY 323 PSLKPEKKVAAEKKVVEBAKKAADQKEDERNPTNTYKTLLEIAESDVEVKAELEL 382  
DB 302 PSLKPEKKVAAEKKVVEBAKKAADQKEDERNPTNTYKTLLEIAESDVEVKAELEL 361  
QY 383 VKEAKEPRNEKYQAKAEVSKKAATRLKIKTDKKAEEBAKKAEDKVEKPA 442  
DB 362 VKEAKEPRNEKYQAKAEVSKKAATRLKIKTDKKAEEBAKKAEDKVEKPA 421  
QY 443 EQPQAPAPAKAEKPAPKPPENPAEQPKAEKPADQAE 481  
DB 422 EQPQAPAPAKAEKPAPKPPENPAEQPKAEKPADQAE 460

RESULT 7  
AA432189  
ID AA432189 standard; Protein; 460 AA.  
AC AA432189;  
XX  
XX 01-FEB-2000 (first entry)  
DT  
DE N-terminal choline binding protein A (Cbpa) truncate.  
XX  
XX Choline binding protein; Cbpa; truncate; adhesin; immunogen;  
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
KW otitis media; pneumonia.  
XX  
XX Streptococcus pneumoniae.  
OS Synthetic.  
XX  
XX WO9951188-A2.  
XX  
PD 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US07669.  
PF  
XX 07-APR-1998; 98US-0056019.  
PR 07-APR-1998; 98US-0080878.  
XX  
XX (SUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;  
PI WPI; 1999-633690/54.  
XX  
DR WPI; 1999-633690/54.  
XX  
XX New N-terminal choline binding protein A truncate polypeptides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections  
XX  
PS Claim 47; Fig 2A-B; 160pp; English.  
XX  
XX The present sequence represents an N-terminal choline binding

CC protein A (Cbpa) truncate, denoted Ntype 4 Cbpa trunc, derived from  
CC Streptococcus pneumoniae Cbpa. Claimed vaccines contain and  
CC N-terminal Cbpa truncate such as the present sequence, or a  
CC polypeptide comprising a conserved region of the Cbpa truncate.  
CC The vaccines provide protection or elicit protective antibodies to  
CC bacterial infection, specifically pneumococcus. Antibodies and  
CC and antagonists against the N-terminal Cbpa truncates are used in  
CC diagnosis and passive immunotherapy.

XX Sequence 460 AA;

Query Match 88.7%; Score 2249.5; DB 20; Length 460;  
Best Local Similarity 98.3%; Pred. No. 2.1e-135;  
Matches 456; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGPKLDSERDKARKEVEEYVK 83  
DB 2 IVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGPKLDSERDKARKEVEEYVK 61  
QY 84 IVGESYAKSTKKRHTITVALVNLNNIKVEYLNKIVESTSESQOIIMMESRSKYDEAVS 143  
DB 62 IVGESYAKSTKKRHTITVALVNLNNIKVEYLNKIVESTSESQOIIMMESRSKYDEAVS 121  
QY 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRN 202  
DB 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRN 181  
QY 203 YPTITTYKTLLEIAESDVEVKAELELVKVKANPEPDQKIQAQAEVSKQAETRLKK 262  
DB 182 YPTITTYKTLLEIAESDVEVKAELELVKVKANPEPDQKIQAQAEVSKQAETRLKK 241  
QY 263 IKTDEEAEAEKRRADAKQGGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 322  
DB 242 IKTDEEAEAEKRRADAKQGGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 301  
QY 323 PSLKPEKKVAEAEKVEEAKKADQKEEDRNYPNTTYKTLLEIAESDVEVKAELEL 382  
DB 302 PSLKPEKKVAEAEKVEEAKKADQKEEDRNYPNTTYKTLLEIAESDVEVKAELEL 361  
QY 383 VKEAEKPEPNEEKVQAQAEVSKQAETRLKIKITDRKKAEEAKRKAEDKYKEKPA 442  
DB 362 VKEAEKPEPNEEKVQAQAEVSKQAETRLKIKITDRKKAEEAKRKAEDKYKEKPA 421  
QY 443 EQOPAPAPKAKPAPKPPENPAQOPKAEKPADQAAE 481  
DB 422 EQOPAPAPKAKPAPKPPENPAQOPKAEKPADQAAE 460

RESULT 8  
AAV49251

ID AAV49251 standard; Protein; 459 AA.

XX AAV49251;

DT 07-FEB-2000 (first entry)

DE N-terminal region of Cbpa polypeptide ATCC4.

KM Choline binding protein A; Cbpa; truncate; immune response; infection;  
XX pneumococcal bacterium; vaccine.

OS Streptococcus pneumoniae.

PN W09951187-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99MO-US07668.

PR 07-APR-1998; 98US-0056019.

PR 07-APR-1998; 98US-0080878.

PA (SUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Tuomanen EI, Maure HR.

DR WPI, 1999-620161/53.

PT Novel polypeptides, used to develop products for the diagnosis,

PT prevention and treatment of pneumococcal infections -

PS Disclosure: Fig 2A-B; 85pp; English.

CC The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC AAV49225, AAV49227, AAV49230, AAV49231, AAV49233 and AAV49238. The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium.  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use.

XX Sequence 459 AA;

Query Match 88.1%; Score 2235.5; DB 20; Length 459;  
Best Local Similarity 98.7%; Pred. No. 1.7e-114;  
Matches 453; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 24 VVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGPKLDSERDKARKEVEEYVK 83  
DB 1 IVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGPKLDSERDKARKEVEEYVK 60  
QY 84 IVGESYAKSTKKRHTITVALVNLNNIKVEYLNKIVESTSESQOIIMMESRSKYDEAVS 143  
DB 61 IVGESYAKSTKKRHTITVALVNLNNIKVEYLNKIVESTSESQOIIMMESRSKYDEAVS 120  
QY 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRN 202  
DB 121 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRN 180  
QY 203 YPTITTYKTLLEIAESDVEVKAELELVKVKANPEPDQKIQAQAEVSKQAETRLKK 262  
DB 181 YPTITTYKTLLEIAESDVEVKAELELVKVKANPEPDQKIQAQAEVSKQAETRLKK 240  
QY 263 IKTDEEAEAEKRRADAKQGGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 322  
DB 241 IKTDEEAEAEKRRADAKQGGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 300  
QY 323 PSLKPEKKVAEAEKVEEAKKADQKEEDRNYPNTTYKTLLEIAESDVEVKAELEL 382  
DB 301 PSLKPEKKVAEAEKVEEAKKADQKEEDRNYPNTTYKTLLEIAESDVEVKAELEL 360  
QY 383 VKEAEKPEPNEEKVQAQAEVSKQAETRLKIKITDRKKAEEAKRKAEDKYKEKPA 442  
DB 361 VKEAEKPEPNEEKVQAQAEVSKQAETRLKIKITDRKKAEEAKRKAEDKYKEKPA 420  
QY 443 EQOPAPAPKAKPAPKPPENPAQOPKAEKPADQAAE 481  
DB 421 EQOPAPAPKAKPAPKPPENPAQOPKAEKPADQAAE 459

RESULT 9

ID AAV32190 standard; Protein; 459 AA.

XX AAV32190;

DT	01-FEB-2000	(first entry)
XX	N-terminal choline binding protein A (Cbpa) truncate.	
DE	Choline binding protein; Cbpa; truncate; adhesin; immunogen; vaccine; diagnosis; therapy: pneumococcus; sepsis; meningitis; otitis media; pneumonia.	
KW	Streptococcus pneumoniae.	
KW	Synthetic.	
XX	MO9951188-A2.	
XX	14-OCT-1999.	
PD	07-APR-1999; 99WO-US07669.	
XX	07-APR-1998; 98US-0056019.	
PR	07-APR-1998; 98US-0080876.	
XX	(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
PA	(MEDI-) MEDIMUNE INC.	
PI	Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;	
XX	WPI: 1999-633690/54.	
DR	New N-terminal choline binding protein A truncate polypeptides, used to develop products for the diagnosis, prevention and treatment of pneumococcal infections -	
PT	Claim 47; Fig 2A-B; 160bp; English.	
XX	The present sequence represents an N-terminal choline binding protein A (Cbpa) truncate, denoted ATCC4 Cbpa trun, derived from Streptococcus pneumoniae Cbpa. Claimed vaccines contain and	
CC	N-terminal Cbpa truncate such as the present sequence, or a polypeptide comprising a conserved region of the Cbpa truncate.	
CC	The vaccines provide protection or elicit protective antibodies to bacterial infection, specifically pneumococcus. Antibodies and	
CC	antagonists against the N-terminal Cbpa truncates are used in diagnosis and passive immunotherapy.	
CC	Sequence 459 AA;	
SO	Query Match 88.1%; Score 2235.5; DB 20; Length 459; Best Local Similarity 98.7%; Pred. No. 1.7e-134; Matches 453; Conservativity 2; Mismatches 3; Indels 1; Gaps 1;	
QY	24 VVASLWMSVVAHATENEGATOVPPSSNNANFSQAEQOGPKLDSERKAREVEEYVK 83	
Db	1 IVASLWMSVVAHATENEGATOVPPSSNNANFSQAEQOGPKLDSERKAREVEEYVK 60	
QY	84 IVGSYAKSTKKRHTTITVALVNLNNINRNEYLNKIVESTSESQOLIMMESRSKYDAVS 143	
Db	61 IVGSYAKSTKKRHTTITVALVNLNNINRNEYLNKIVESTSESQOLIMMESRSKYDAVS 120	
QY	144 KFEKDSSSSSSDSTKPEASDTAKPNKPTBEGEYVAEAKKKVEEA-KYAKDQKEEDRN 202	
Db	121 KFEKDSSSSSSDSTKPEASDTAKPNKPTBEGEYVAEAKKKVEEYKAKDQKEEDRN 180	
QY	203 YPTITTYKTLLEIASDVEYKKAELVKKVANRPRDQKI KOAEAFVESKQAEYTRKK 262	
Db	181 YPTITTYKTLLEIASDVEYKKAELVKKVANRPRDQKI KOAEAFVESKQAEYTRKK 240	
QY	263 IKTDREAEAEBAKRRADAKQKPKGRKRGVPGELATPDKKENDAKSDSGVGETLPS 322	
Db	241 IKTDREAEAEBAKRRADAKQKPKGRKRGVPGELATPDKKENDAKSDSGVGETLPS 300	
QY	323 PSLPDEKKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEYKKAEL 382	
Db	301 PSLPDEKKVAEAEKKVEAKKKAEDQKEDRRNPTNTYKTLLEIASDVEYKKAEL 360	

QY		363	VKEAKEPERNEKYNQAALAEVSESKAENATRLREKITDRKKAEEAKRKAAEDPKYKEXRA	442
Db		361	VKEEAKERNEKYNQAARAEVSESKAEATRLERKITDRKKAAEEAKRKAAEDPKYKEXRA	420
QY		443	EOPAPAPAKAEKPAPAKPEPENPAPQPAKAEKPADDOAAE	481
Db		421	EOPAPAPAKTEKPPAPAKPEPENPAPQPAKAEKPADDOAAE	459
<hr/>				
	RESULT 10			
ID	AA49143			
XX	AA49143 standard; Protein; 446 AA.			
AC				
XX	AA49143;			
XX				
DT	17-JAN-2000 (first entry)			
XX				
DE	Amino acid sequence of choline-binding protein fragment #7.			
KW	Truncated surface binding protein; alpha helix; choline binding protein;			
KM	vaccinia; invasive bacterial infection; otitis media; sepsis;			
KW	meningitis; lobar pneumonia infection; antibody; immature immune system;			
KX	immunocompromised.			
XX				
OS	Streptococcus pneumoniae.			
XX				
FN	WO9951266-A2.			
PD				
XX	14-OCT-1999.			
PF				
XX	06-APR-1999; 99WO-US07680.			
PR				
XX	07-APR-1998; 98US-0080878.			
PR				
XX	15-MAY-1998; 98US-0085743.			
PA	(MED1-) MEDIMUNE INC.			
P1				
P1	Wizemann TM, Koenig S, Johnson LS;			
DR	WPI; 1999-601465/51.			
DR	N-PDSB; AAZ31083.			
PT	New pneumococcal proteins useful as vaccines and for diagnosis of			
PT	pneumococcal infections -			
PS	Claim 10; Page 72-73; 98pp; English.			
XX				
CC	AA49137-V49152 are amino acid sequences that are fragments of choline			
CC	binding proteins (CBP). The fragments of the protein are the alpha helix			
CC	forming parts of the CBPs from Streptococcus pneumoniae. The			
CC	polypeptides do not contain the actual choline binding fragment. The			
CC	polypeptides and the nucleotide sequences that encode them			
CC	(AAZ31077-ZJ31092) are used in the invention, which relates to polypeptide			
CC	truncates of a pneumococcal surface binding protein containing the highly			
CC	conserved immunogenic alpha helical portion and no choline binding			
CC	portion. The polypeptides are used as immunogens in a bacterial vaccine.			
CC	bacterial (especially pneumococcal) infections, especially otitis media			
CC	(caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia			
CC	infections. Antibodies raised against the polypeptide are useful for			
CC	detection, prevention (passive immunity) and treatment of S. pneumoniae			
CC	infections. The vaccines are especially useful in immunocompromised			
CC	patients, those with an immature immune system, or patients with an on			
CC	going pneumococcal infection. The vaccine avoids unnecessary expense and			
CC	provides broad protection against a range of pneumococcal serotypes and			
CC	it produces an improved and enhanced effect in preventing bacterial			
CC	infections.			
SQ				
Sequence	446 AA;			
Query Match	86.5%; Score 2194.5; DB 20; Length 446;			
Best Local Similarity	99.8%; Pred. No. 6.5e-132;			
Matches 446; Conservative	0; Mismatches 0; Indels 1; Gaps 1			

```
QY 37 TENEGATOVPTSSNRANESQAEQGPCKLDSEBDKARKEVEYVKTIIGSYAKSTKKR 96
DB 1 TENEGATOVPTSSNRANESQAEQGPCKLDSEBDKARKEVEYVKTIIGSYAKSTKKR 60
QY 97 HTITVALVNLINIKNEYLINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSD 156
DB 61 HTITVALVNLINIKNEYLINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSD 120
QY 157 SSTKEPADTAKPKNPTPEGKVAEAKKKVBEA-KKADQKEDRRNPTITTYKTLELEI 215
DB 121 SSTKEPADTAKPKNPTPEGKVAEAKKKVBEA-KKADQKEDRRNPTITTYKTLELEI 180
QY 216 AESDVEVKKALELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEBAK 275
DB 181 AESDVEVKKALELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEBAK 240
QY 276 RRAADAKKQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEPKVAEAE 335
DB 241 RRAADAKKQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEPKVAEAE 300
QY 336 KKVBEAKKKAADQKEDRRNPTITTYKTLELEIASDVEVKKALELVKEAEKPEPRNEEK 395
DB 301 KKVBEAKKKAADQKEDRRNPTITTYKTLELEIASDVEVKKALELVKEAEKPEPRNEEK 360
QY 336 VKQAKAEVSKKAETRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 455
DB 361 VKQAKAEVSKKAETRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPNPAPQPAEKPADQQAEE 481
DB 421 PAPAPKPNPAPQPAEKPADQQAEE 446
RESULT 11
ID AAY49140
AAAY49140 standard; Protein; 446 AA.
AC AAY49140;
DE 17-JAN-2000 (first entry)
XX Amino acid sequence of choline-binding protein fragment #4.
XX Truncated surface binding protein; alpha helix; choline binding protein;
XX vaccine; invasive bacterial infection; otitis media; sepsis;
XX meningitis; lobar pneumonia infection; antibody; immature immune system;
XX immunocompromised.
OS Streptococcus pneumoniae.
XX Key Location/Qualifiers
FH Misc-difference 171 /note= "Encoded by AAT"
FT Misc-difference 172 /note= "Encoded by TAC"
FT Misc-difference 173 /note= "Encoded by TAA"
FT Misc-difference 174 /note= "Encoded by TTA"
FT Misc-difference 174 /note= "Encoded by CAA"
XX MO9951266-A2.
XX 14-OCT-1999.
XX 06-APR-1999; 99WO-US07680.
XX 07-APR-1998; 98US-0080878.
XX 15-MAY-1998; 98US-0085743.
XX (MEDI-) MEDIMUNE INC.
XX Wizemann TM, Koenig S, Johnson LS;
```

```
XX WPI: 1999-601465/51.
DR N-PSDB; AA231080.
XX New pneumococcal proteins useful as vaccines and for diagnosis of
PT pneumococcal infections -
PS Claim 10; Page 68-69; 98pp; English.
XX AAAY9137-Y49152 are amino acid sequences that are fragments of choline
CC binding proteins (CBP). The fragments of the protein are the alpha helix
CC forming parts of the CBPs from Streptococcus pneumoniae. The
CC polypeptides do not contain the actual choline binding fragment. The
CC polypeptides and the nucleotide sequences that encode them
CC (AA231077-231092) are used in the invention, which relates to polypeptide
CC truncates of a pneumococcal surface binding protein containing the highly
CC conserved immunogenic alpha helical portion and no choline binding
CC portion. The polypeptides are used as immunogens in a bacterial vaccine.
CC The vaccine can be used for preventing (immunizing) or treating invasive
CC bacterial (especially pneumococcal) infections, especially otitis media
CC (caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia
CC infection. Antibodies raised against the polypeptide are useful for
CC detection, prevention (passive immunity) and treatment of S. pneumoniae
CC infections. The vaccines are especially useful in immunocompromised
CC patients, those with an immature immune system, or patients with an on
CC going pneumococcal infection. The vaccine avoids unnecessary expense and
CC it provides broad protection against a range of pneumococcal serotypes and
CC it produces an improved and enhanced effect in preventing bacterial
CC infections.
SQ Sequence 446 AA;
```

Query Match 85.8%; Score 2177.5; DB 20; Length 446;

Best Local Similarity 98.9%; Pred. No. 7.9e-131;

Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

```
QY 37 TENEGATOVPTSSNRANESQAEQGPCKLDSEBDKARKEVEYVKTIIGSYAKSTKKR 96
DB 1 TENEGATOVPTSSNRANESQAEQGPCKLDSEBDKARKEVEYVKTIIGSYAKSTKKR 60
QY 97 HTITVALVNLINIKNEYLINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSD 156
DB 61 HTITVALVNLINIKNEYLINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSD 120
QY 157 SSTKEPADTAKPKNPTPEGKVAEAKKKVBEA-KKADQKEDRRNPTITTYKTLELEI 215
DB 121 SSTKEPADTAKPKNPTPEGKVAEAKKKVBEA-KKADQKEDRRNPTITTYKTLELEI 180
QY 216 AESDVEVKKALELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEBAK 275
DB 181 AESDVEVKKALELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEBAK 240
QY 276 RRAADAKKQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEPKVAEAE 335
DB 241 RRAADAKKQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEPKVAEAE 300
QY 336 KKVBEAKKKAADQKEDRRNPTITTYKTLELEIASDVEVKKALELVKEAEKPEPRNEEK 395
DB 301 KKVBEAKKKAADQKEDRRNPTITTYKTLELEIASDVEVKKALELVKEAEKPEPRNEEK 360
QY 336 VKQAKAEVSKKAETRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 455
DB 361 VKQAKAEVSKKAETRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPNPAPQPAEKPADQQAEE 481
DB 421 PAPAPKPNPAPQPAEKPADQQAEE 446
RESULT 12
ID AAY49238
AAAY49238 standard; Protein; 428 AA.
```





Matches 427; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97  
 DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60  
 QY 98 TITVALVNLNNIKNEYLINKIVESTSESQOILMMESRSKYDEAVSKFEKSSSSSSDS 157  
 DB 61 TITVALVNLNNIKNEYLINKIVESTSESQOILMMESRSKYDEAVSKFEKSSSSSSDS 120  
 QY 158 STKPEASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 216  
 DB 121 STKPEASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 180  
 QY 217 ESDVVKKALELVVKANEPDRDQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 276  
 DB 181 ESDVVKKALELVVKANEPDRDQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 240  
 QY 277 RADAKEQCKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAEK 336  
 DB 241 RADAKEQCKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAEK 300  
 QY 337 KVEEAKKKAEDQKEDRRNYPITYKTLELEIAESDVEYKKALELVKEAEKPRNEEKV 396  
 DB 301 KVEEAKKKAEDQKEDRRNYPITYKTLELEIAESDVEYKKALELVKEAEKPRNEEKV 360  
 QY 337 KQAKAEVSKKAATRLKIKITDRKKAEEAKKAEAEKVAEAEQOPAPAKAEKP 456  
 DB 361 KQAKAEVSKKAATRLKIKITDRKKAEEAKKAEAEKVAEAEQOPAPAKAEKP 420  
 QY 457 APAPKREN 464  
 DB 421 APAPKREN 428

RESULT 14  
 AAY49225  
 ID AAY49225 standard; protein; 406 AA.  
 XX AAY49225;  
 AC AAY49225;  
 DT 07-FEB-2000 (first entry)  
 XX  
 DE Polypeptide R2 of the N-terminal truncate of Cbpa serotype type 4.  
 XX  
 KW Choline binding protein A; Cbpa; truncate; immune response; infection;  
 KM pneumococcal bacterium; vaccine.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9951187-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-US07668.  
 XX  
 PR 07-APR-1998; 98US-0056019.  
 PR 07-APR-1998; 98US-0080878.  
 XX  
 PA (SUJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 XX  
 PI Tuomanen EI, Maure HR;  
 XX  
 DR WPI; 1999-620161/53.  
 DR N-PSDB; AA231401.  
 XX  
 PT Novel polypeptides, used to develop products for the diagnosis,  
 PT prevention and treatment of pneumococcal infections  
 XX  
 PS Claim 2; Page 7-8; 85pp; English.  
 PS  
 CC The invention provides novel isolated polypeptides comprising the amino  
 CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate

CC (Cbpa). The polypeptides can be selected from sequences shown in  
 CC AAY49225, AAY49227, AAY49230, AAY49231, AAY49233 and AAY49238. The  
 CC polypeptides can be used for inducing an immune response in a subject  
 CC which has been exposed to or infected with a pneumococcal bacterium. They  
 CC can also be used for preventing infection by a pneumococcal bacterium.  
 CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
 CC used for treating a subject infected with or exposed to a pneumococcal  
 CC bacterium. Antibodies specifically binding the polypeptides can be used  
 CC for detection and diagnosis and for preventing pneumococcal attachment to  
 CC a mucosal surface. The products can be used in humans and other animals  
 CC such as domestic animals, such as feline or canine subjects, farm animals  
 CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
 CC animals (whether in the wild or in a zoological garden), research  
 CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
 CC i.e. for veterinary medical use.

XX  
 SQ Sequence 406 AA;

Query Match 78.0%; Score 1979.5; DB 20; Length 406;  
 Best Local Similarity 99.8%; Pred. No. 2.9e-118;  
 Matches 405; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97  
 DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60  
 QY 98 TITVALVNLNNIKNEYLINKIVESTSESQOILMMESRSKYDEAVSKFEKSSSSSSDS 157  
 DB 61 TITVALVNLNNIKNEYLINKIVESTSESQOILMMESRSKYDEAVSKFEKSSSSSSDS 120  
 QY 158 STKPEASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 216  
 DB 121 STKPEASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 180  
 QY 217 ESDVVKKALELVVKANEPDRDQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 276  
 DB 181 ESDVVKKALELVVKANEPDRDQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 240  
 QY 277 RADAKEQCKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAEK 336  
 DB 241 RADAKEQCKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAEK 300  
 QY 337 KVEEAKKKAEDQKEDRRNYPITYKTLELEIAESDVEYKKALELVKEAEKPRNEEKV 396  
 DB 301 KVEEAKKKAEDQKEDRRNYPITYKTLELEIAESDVEYKKALELVKEAEKPRNEEKV 360  
 QY 397 KQAKAEVSKKAATRLKIKITDRKKAEEAKKAEAEKVAEAEQOPAPAKAEKP 442  
 DB 361 KQAKAEVSKKAATRLKIKITDRKKAEEAKKAEAEKVAEAEQOPAPAKAEKP 406

RESULT 15  
 AAY32098  
 ID AAY32098 standard; Protein; 406 AA.  
 XX AAY32098;  
 AC AAY32098;  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Choline binding protein A (Cbpa) truncate R2.  
 XX  
 KW Choline binding protein; Cbpa; truncate; adhesion; immunogen;  
 KM vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9951188-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PR 07-APR-1999; 99WO-US07669.





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:02 ; Search time 87.0043 Seconds  
(without alignments)  
1882.497 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442  
Perfect score: 889  
Sequence: 1 IKTDREAESEBAKRRADAKE.....AEEBAKKAEDKVKKEKPA 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Pending Patents AA Main:\*

- 1: /cgn2\_6/ptodata/1/paa/PCRRUS\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/1/paa/US097A\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/1/paa/US097B\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/1/paa/US099A\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/1/paa/US099B\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*
- 28: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*
- 29: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*
- 30: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*
- 31: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep.\*
- 32: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	406	14	US-09-056-019-1 Sequence 1, Appl

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45																																																																																																																																																																																																										
889	100.0	406	14	US-09-056-019B-1	Sequence 1, Appl	889	100.0	406	14	US-09-287-788-1	Sequence 1, Appl	889	100.0	428	14	US-09-056-019-24	Sequence 24, Appl	889	100.0	428	14	US-09-056-019B-24	Sequence 24, Appl	889	100.0	428	16	US-09-287-788-24	Sequence 24, Appl	889	100.0	446	28	US-10-254-995-9	Sequence 9, Appl	889	100.0	453	13	US-08-961-083-38	Sequence 38, Appl	889	100.0	453	22	US-09-765-271-38	Sequence 38, Appl	889	100.0	453	22	US-09-765-272-38	Sequence 38, Appl	889	100.0	453	22	US-09-765-272A-38	Sequence 38, Appl	889	100.0	460	14	US-09-056-019-38	Sequence 38, Appl	889	100.0	460	14	US-09-056-019B-38	Sequence 38, Appl	889	100.0	460	16	US-09-287-788-38	Sequence 38, Appl	889	100.0	511	16	US-09-298-523B-13	Sequence 13, Appl	889	100.0	511	16	US-09-298-523B-13	Sequence 13, Appl	889	100.0	511	21	US-09-748-875-13	Sequence 13, Appl	889	100.0	511	29	US-10-341-201-13	Sequence 13, Appl	889	100.0	513	16	US-09-298-523B-12	Sequence 12, Appl	889	100.0	513	16	US-09-298-523C-12	Sequence 12, Appl	889	100.0	513	21	US-09-748-875-12	Sequence 12, Appl	889	100.0	513	29	US-10-341-201-12	Sequence 12, Appl	889	100.0	653	14	US-09-056-019-2	Sequence 2, Appl	889	100.0	655	14	US-09-056-019B-2	Sequence 2, Appl	889	100.0	655	16	US-09-287-788-2	Sequence 2, Appl	889	100.0	693	22	US-09-769-787-185	Sequence 185, App	889	100.0	693	29	US-10-333-120A-10	Sequence 10, Appl	884	99.4	446	28	US-10-254-995-6	Sequence 6, Appl	884	99.4	459	14	US-09-056-019-39	Sequence 39, Appl	884	99.4	459	16	US-09-056-019B-39	Sequence 39, Appl	884	99.4	459	16	US-09-287-788-39	Sequence 39, Appl	861	96.9	451	16	US-09-298-523B-67	Sequence 67, Appl	861	96.9	451	21	US-09-748-875-67	Sequence 67, Appl	861	96.9	451	29	US-10-341-201-67	Sequence 67, Appl	856	96.3	437	14	US-09-056-019B-40	Sequence 40, Appl	803.5	90.4	376	14	US-09-056-019-7	Sequence 7, Appl	803.5	90.4	376	16	US-09-287-788-7	Sequence 7, Appl	803.5	90.4	413	14	US-09-056-019-35	Sequence 35, Appl	803.5	90.4	413	16	US-09-287-788-35	Sequence 35, Appl	803.5	90.4	413	28	US-10-254-995-16	Sequence 16, Appl	803.5	90.4	581	16	US-09-298-523B-56	Sequence 56, Appl	803.5	90.4	581	16	US-09-298-523C-56	Sequence 56, Appl

#### ALIGNMENTS

RESULT 1  
US-09-056-019-1  
Sequence 1, Application US/09056019A  
GENERAL INFORMATION:  
APPLICANT: Tuomenen, Elaine I  
APPLICANT: Mitzemann, Theresa  
APPLICANT: Maesure, H. R.  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLESTEROL BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056, 019A  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-1

Query Match 100.0%; Score 889; DB 14; Length 406;

Best Local Similarity 100.0%; Pred. No. 1.5e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 120  
DB 287 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 346  
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 2  
US-09-056-019B-1  
; Sequence 1, Application US/09056019B  
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzemann, Theresa M.  
APPLICANT: Masure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 5853-2  
CURRENT APPLICATION NUMBER: US/09/056,019B  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019B-1

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.5e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 120  
DB 287 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 346  
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 3  
US-09-287-788-1  
; Sequence 1, Application US/09287788B  
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I  
APPLICANT: Masure, H. R.  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017N  
CURRENT APPLICATION NUMBER: US/09/287,788B  
CURRENT FILING DATE: 1998-04-07  
PRIOR APPLICATION NUMBER: 60/080,878  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1

LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-287-788-1

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.5e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 120  
DB 287 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 346  
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 4  
US-09-056-019-24  
; Sequence 24, Application US/09056019A  
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzemann, Theresa  
APPLICANT: Masure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056,019A  
CURRENT FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-24

Query Match  
Best Local Similarity 100.0%; Pred. No. 1.6e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 120  
DB 287 PSLEPEKVAEAEKKEVEAKKKAEDQKEDRRNYPNTYKTLLELEIAESDVEVKKALEL 346  
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAKEPRNEEKVQKAAEVESKKAATRLKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 5  
US-09-056-019B-24

; Sequence 24, Application US/09056019B  
; GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzemann, Theresa M.  
APPLICANT: Masure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED

;; TITLE OF INVENTION: THEREFROM AND USES THEREOF  
;; FILE REFERENCE: 5853-2  
;; CURRENT APPLICATION NUMBER: US/09/056,019B  
;; CURRENT FILING DATE: 1998-04-07  
;; NUMBER OF SEQ ID NOS: 40  
;; SOFTWARE: Patent In Ver. 2.0  
;; SEQ ID NO 24  
;; LENGTH: 428  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
US-09-056-019B-24

Query Match 100.0%; Score 889; DB 14; Length 428;  
Best Local Similarity 100.0%; Pred. No. 1.6e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 227 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 286  
Qy 61 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
Db 287 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346  
Qy 121 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 347 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 6  
US-09-287-788-24  
Sequence 24, Application US/09287788B  
GENERAL INFORMATION:  
APPLICANT: Tuomenen, Elaine I  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
TITLE OF INVENTION: THEREFROM AND USES THEREOF  
FILE REFERENCE: 1340-1-017N  
CURRENT APPLICATION NUMBER: US/09/287,788B  
CURRENT FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: 60/080,878  
PRIOR FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 24  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-287-788-24

Query Match 100.0%; Score 889; DB 16; Length 428;  
Best Local Similarity 100.0%; Pred. No. 1.6e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 227 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 286  
Qy 61 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
Db 287 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346  
Qy 121 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 347 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 7  
US-10-254-995-9  
Sequence 9, Application US/10254995  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.

;; APPLICANT: Koenig, Scott  
;; APPLICANT: Johnson, Leslie S  
;; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
;; FILE REFERENCE: 469201-396  
;; CURRENT APPLICATION NUMBER: US/10/254,995  
;; CURRENT FILING DATE: 2002-09-25  
;; PRIOR APPLICATION NUMBER: US/09/286,981  
;; PRIOR FILING DATE: 1999-04-06  
;; PRIOR APPLICATION NUMBER: US 60/085,743  
;; PRIOR FILING DATE: 1998-05-15  
;; NUMBER OF SEQ ID NOS: 38  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 9  
;; LENGTH: 446  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
US-10-254-995-9

Query Match 100.0%; Score 889; DB 28; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
Db 228 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 287  
Qy 61 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
Db 288 PSLEPEKVAEAEKVEAEKKAADOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 347  
Qy 121 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
Db 348 VKEAKEPRNEEKYKQAKAEVESKKAETRLKIKTRKKAEEBAKRAAEEDVKKEKPA 407

RESULT 8  
US-08-961-083-38  
Sequence 38, Application US/08961083  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: <Unknown>  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-08-961-083-38

Query Match 100.0%; Score 889; DB 13; Length 453;

Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 60  
DB 25 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 84  
QY 61 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
DB 85 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144  
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 180  
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 204

RESULT 9  
US-09-765-271-38

Sequence 38, Application US/09765271

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,271

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/536,784

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 453 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-09-765-271-38

Query Match 100.0%; Score 889; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 60

DB 25 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 84  
QY 61 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
DB 85 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144  
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 180  
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 204

RESULT 10  
US-09-765-272-38

Sequence 38, Application US/09765272

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brooks, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 453 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-09-765-272-38

Query Match 100.0%; Score 889; DB 22; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 60  
DB 25 IKTDREAEAEAKRADAKGKPKGRAGVPGELATPPKENDAKSSDSSVGEETLPS 84  
QY 61 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120  
DB 85 PSLEKPKVAEAEKVEAEKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144  
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 180  
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTRKKAEBAKRAAEEDVKEXKPA 204

RESULT 11  
US-09-765-272A-38

```
; Sequence 38, Application US/09765272A
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and
;   Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
;   ADDRESS: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
;   COMPUTER: Dell Latitude C610
;   OPERATING SYSTEM: Windows 2000
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/765,272A
;     FILING DATE: 22-Jan-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/961,083
;     FILING DATE: OCT-30-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Lin J. Hymel
;     REGISTRATION NUMBER: 45,414
;     REFERENCE/DOCKET NUMBER: PB340P2C2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (301) 610-5790
;     TELEFAX: (301) 309-8439
;   INFORMATION FOR SEQ ID NO: 38:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 453 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-765-272A-38

Query Match          100.0%; Score 889; DB 22; Length 453;
Best Local Similarity 100.0%; Pred. No. 1,7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERAKRADAKGKPKRAKRGVPELATPPDKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREAEAEERAKRADAKGKPKRAKRGVPELATPPDKENDAKSSDSSVGEETLPS 84
QY 61 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKKALEL 120
DB 85 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKKALEL 144
QY 121 VKERAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 145 VKERAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 204

RESULT 12
US-09-056-019-38
; Sequence 38, Application US/09056019A
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wisemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
```

```
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match          100.0%; Score 889; DB 14; Length 460;
Best Local Similarity 100.0%; Pred. No. 1,7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERAKRADAKGKPKRAKRGVPELATPPDKENDAKSSDSSVGEETLPS 60
DB 242 IKTDREAEAEERAKRADAKGKPKRAKRGVPELATPPDKENDAKSSDSSVGEETLPS 301
QY 61 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKKALEL 120
DB 302 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKKALEL 361
QY 121 VKERAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 362 VKERAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 421

RESULT 14
US-09-287-788-38
; Sequence 38, Application US/09287788B
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Masure, H. R.
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
```

;; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
;; TITLE OF INVENTION: THEREFROM AND USES THEREOF  
;; FILE REFERENCE: 1340-1-017N  
;; CURRENT APPLICATION NUMBER: US/09/287,788B  
;; PRIOR FILING DATE: 1999-04-07  
;; PRIOR APPLICATION NUMBER: 60/080,878  
;; PRIOR FILING DATE: 1998-04-07  
;; NUMBER OF SEQ ID NOS: 39  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 38  
;; LENGTH: 460  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
US-09-287-788-38

Query Match 100.0%; Score 889; DB 16; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.7e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAEKRRADAEQGGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
DB 242 IKTDREAEAEAEKRRADAEQGGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPS 301  
QY 61 PSLEPEKKVAEAEKKVEAEKKAEDQKEDRRNYPTNTYKTLLEIAESDVEVKKALEL 120  
DB 302 PSLEPEKKVAEAEKKVEAEKKAEDQKEDRRNYPTNTYKTLLEIAESDVEVKKALEL 361  
QY 121 VKEAKEPRNEBKVKQAAEVESSKAAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 180  
DB 362 VKEAKEPRNEBKVKQAAEVESSKAAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 421

## RESULT 15

US-09-298-523B-13  
; Sequence 13, Application US/09298523B  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-13

Query Match 100.0%; Score 889; DB 16; Length 511;  
Best Local Similarity 100.0%; Pred. No. 2e-49;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAEKRRADAEQGGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
DB 263 IKTDREAEAEAEKRRADAEQGGKPKRAKRGVPGELATPDKKENDAKSSDSSVGEETLPS 322  
QY 61 PSLEPEKKVAEAEKKVEAEKKAEDQKEDRRNYPTNTYKTLLEIAESDVEVKKALEL 120  
DB 323 PSLEPEKKVAEAEKKVEAEKKAEDQKEDRRNYPTNTYKTLLEIAESDVEVKKALEL 382  
QY 121 VKEAKEPRNEBKVKQAAEVESSKAAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 180  
DB 383 VKEAKEPRNEBKVKQAAEVESSKAAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 442

Search completed: November 21, 2003, 13:40:58  
Job time : 88.0043 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:37 ; Search time 7.03329 Seconds  
(without alignments)  
1285.946 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442

Perfect score: 889

Sequence: 1 IKTDREAEAEKRRADAKE.....AEEKAKKAEEDKVEKPA 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 261868-seqs, 50246823 residues

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/prodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/prodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	889	100.0	693	1	PCT-US03-27401-468 Sequence 468, App
2	889	100.0	693	6	US-10-472-928-4598 Sequence 4598, App
3	187.5	21.1	1026	6	US-10-258-899A-1487 Sequence 1487, App
4	187.5	21.1	1033	6	US-10-258-899A-3455 Sequence 3455, App
5	162.5	18.3	507	6	US-10-425-114A-47285 Sequence 47285, App
6	158.5	17.8	257	1	PCT-US03-11802-20 Sequence 20, App
7	158.5	17.8	487	1	PCT-US03-11802-21 Sequence 21, App
8	156	17.5	744	1	PCT-US03-27401-251 Sequence 251, App
9	156	17.5	744	6	US-10-472-928-32 Sequence 32, App
10	154.5	17.4	1533	6	US-10-679-063-23311 Sequence 23311, App
11	152	17.1	793	7	US-60-490-890-2605 Sequence 2605, App
12	148	16.6	725	6	US-10-258-899A-1881 Sequence 1881, App
13	148	16.6	774	6	US-10-258-899A-3849 Sequence 3849, App
14	146	16.4	369	5	US-09-820-843B-95 Sequence 95, App
15	146	16.4	665	5	US-09-820-843B-107 Sequence 107, App
16	146	16.4	938	7	US-60-490-890-495 Sequence 495, App
17	145.5	16.4	356	5	US-09-820-843B-27 Sequence 27, App
18	143.5	16.1	4684	1	PCT-US03-14382-432 Sequence 432, App
19	143.5	16.1	4684	1	PCT-US03-14382-434 Sequence 434, App
20	143.5	16.1	4684	1	PCT-US03-14382A-432 Sequence 432, App
21	143.5	16.1	4684	1	PCT-US03-14382A-434 Sequence 434, App
22	143.5	16.1	4684	7	US-60-479-073-497 Sequence 497, App
23	143.5	16.1	4684	7	US-60-479-073-506 Sequence 506, App
24	141	15.9	612	7	US-60-500-337-2473 Sequence 2473, App
25	141	15.9	612	7	US-60-500-337-2475 Sequence 2475, App
26	141	15.9	662	7	US-60-500-337-2476 Sequence 2476, App

27	140	15.7	2524	7	US-60-502-656-219 Sequence 219, App
28	140	15.7	2524	7	US-60-512-690-253 Sequence 253, App
29	140	15.7	4574	7	US-60-479-073-495 Sequence 495, App
30	140	15.7	4576	1	PCT-US03-3147C-281 Sequence 281, App
31	139.5	15.7	916	7	US-60-490-890-475 Sequence 475, App
32	138.5	15.6	372	5	US-09-820-843B-8 Sequence 8, App
33	138	15.5	204	1	PCT-US03-26864-304 Sequence 304, App
34	138	15.5	204	1	PCT-US03-27107-304 Sequence 304, App
35	138	15.5	572	6	US-10-425-114A-43212 Sequence 43212, App
36	137.5	15.5	486	6	US-10-679-063-14456 Sequence 14456, App
37	137.5	15.5	543	6	US-10-679-063-14127 Sequence 14127, App
38	137.5	15.5	926	5	US-09-897-516A-6173 Sequence 6173, App
39	137	15.4	558	7	US-60-490-890-2607 Sequence 2607, App
40	137	15.4	564	7	US-60-490-890-2611 Sequence 2611, App
41	137	15.4	1027	7	US-60-495-114A-1376 Sequence 1376, App
42	136.5	15.4	284	7	US-60-485-450-1572 Sequence 1572, App
43	136.5	15.4	515	6	US-10-343-650A-418 Sequence 418, App
44	136.5	15.4	1307	6	US-10-258-899A-3084 Sequence 3084, App
45	136.5	15.4	1307	6	US-10-258-899A-3085 Sequence 3085, App

#### ALIGNMENTS

```
RESULT 1
PCT-US03-27401-468
; Sequence 468, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; PRIOR FILING DATE: 2003-09-02
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 468
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; PCT-US03-27401-468

Query Match      100.0%; Score 889; DB 1; Length 693;
Best Local Similarity 100.0%; Pred. No. 5.1e-32;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IKTDREAEAEKRRADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
DB      265 IKTDREAEAEKRRADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 324

QY      61 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEYKAELEI 120
DB      325 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEYKAELEI 384

QY      121 VEEAEKPEPNEKVKQAKAEVSKKAEATRLKIKITDRKKAEEAKKAAEDKVEKPA 180
DB      385 VEEAEKPEPNEKVKQAKAEVSKKAEATRLKIKITDRKKAEEAKKAAEDKVEKPA 444

RESULT 2
US-10-472-928-4598
; Sequence 4598, Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
```

```

NUMBER OF SEQ ID NOS: 4979
SOFTWARE: SeqMtn39, version 1.03
SEQ ID NO 4598
LENGTH: 693
TYPE: prt
ORGANISM: Streptococcus pneumoniae
FEATURES:
OTHER INFORMATION: choline binding protein A (cbpa)
OTHER INFORMATION: Cellular location: outside
OTHER INFORMATION: Feature of note: WY motif
OTHER INFORMATION: Similar to strain R6 sequence 15904036 (O.E+01)
US-10-472-928--4598

```

Query Match	100.0%	Score 889	DB 6	Length 693
Best Local Similarity	100.0%	Pred. No. 5.1e-32		
Matches 180	0	Mismatches 0	Indels 0	Gaps 0

Qy	1	KTDEEBAEAEKRRADAEQCKPKRARGVGELATDPDKKENDKSSSVSGEETLP	60
Db	265	IKTDEEBAEAEKRRADAEQCKPKRARGVGELATDPDKKENDKSSSVSGEETLP <td>324</td>	324
Qy	61	PSLKEKKVAAEAKKVEEAKKKAEEDCKEDRRVYPNTYKTTLETLAESDVEVKAELEL	120
Db	325	PSLKEKKVAAEAKKVEEAKKKAEEDCKEDRRVYPNTYKTTLETLAESDVEVKAELEL	384
Qy	121	VKEAEKPEPNEKVKQAKAEVSSKKAETARLEKIKTDKKAEEBAARKAAEDVKVEKPA	180
Db	385	VKEAEKPEPNEKVKQAKAEVSSKKAETARLEKIKTDKKAEEBAARKAAEDVKVEKPA	444

### RESULT 3

```

US-10-258699A-1487
: Sequence 1487, Application US/10258699A
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Dimanac, Radjoe T.
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhou, Ping
: APPLICANT: Xu, Chongjun
: APPLICANT: Cao, Yicheng
: APPLICANT: Ma, Yungqing
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyun
: APPLICANT: Chen, Rui-hong
: APPLICANT: Wang, Zhi Wei
: APPLICANT: Xue, Aidong
: APPLICANT: Yang, Yonghong
: APPLICANT: Weinman, Tom
: APPLICANT: Goodrich, Ryle
: TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
: FILE REFERENCE: 78/CIP2-2G/US
: CURRENT APPLICATION NUMBER: US/10/258, 899A
: CURRENT FILING DATE: 2003-11-06
: PRIOR APPLICATION NUMBER: PCT/US01/04098
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: 09/774, 434
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: 09/728, 422
: PRIOR FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: 09/693, 325
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 09/663, 561
: PRIOR FILING DATE: 2000-09-15
: PRIOR APPLICATION NUMBER: 09/654, 936
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 09/620, 325
: PRIOR FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: 09/598, 075
: PRIOR FILING DATE: 2000-06-20

```

```

; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1487
; LENGTH: 1026
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-258-899A-1487

```

Query Match	21.1%	Score 187.5	DB 6	Length 1026
Best Local Similarity	31.9%	Pred. NO. 0.075		
Matches 60	Conservative 29	Mismatches 76	Indels 23	Gaps 7

```
QY      5 REAEAEARRDADAEQOGPKGRARAGVPGELAPDDKENDAKSDSVEBETLPSPSLK 64
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      705 KEBAKSPPEKASPVKEEAKSPPEKASPVKEAKTPPEKAKSPVKEAKS-----PEKAKS 758
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      65 PEKVAEAERKVEAEAKKKAEDQEDRRNRYPTNTYKLTLEI-----AESDV--EYKAE 117
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      759 PE-KATLDVSKPEAKTPAKEARSPPADKPEKAKSPVKEEIVKSPEKAKSPLEMDAKAPE 817
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      118 LED-VKEAEKPRNEE-----KYQAAVAEVESKKAETRIEIKYDRKKAEEBAKAKAA 170
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      818 KEIPKKEEVKSPVKEEEKQOEIVKAPPEKKAEEERAPATPTEEKDOKSK--EEAPKEA 875
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      171 EEDKVKEK 178
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      876 PKPKVEEK 883
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

## RESULT 4

US-10-258-899A-3455  
Sequence 3455, Application US/10258899A  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Drmanac, Radcoje T.  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhou, Ping  
APPLICANT: Xu, Chongjun  
APPLICANT: Cao, Yicheng  
APPLICANT: Ma, Yanguang  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wang, Duntui  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Wang, Zhi Wei  
APPLICANT: Xue, Aidong  
APPLICANT: Yang, Yonghong  
APPLICANT: Weinman, Tom  
APPLICANT: Goodrich, Ryle  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 78CIPC2-2G/US  
CURRENT APPLICATION NUMBER: US/10/258, 899A  
CURRENT FILING DATE: 2003-11-06  
PRIOR APPLICATION NUMBER: PCT/US01/04098  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 09/774, 434  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 09/728, 422  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: 09/693, 325  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 09/663, 561  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 09/654, 936  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 09/620, 325

;; PRIOR FILING DATE: 2000-07-19  
;; PRIOR APPLICATION NUMBER: 09/598,075  
;; PRIOR FILING DATE: 2000-06-20  
;; PRIOR APPLICATION NUMBER: 09/560,875  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: 09/496,914  
;; PRIOR FILING DATE: 2000-02-03  
;; NUMBER OF SEQ ID NOS: 3960  
;; SOFTWARE: Custom  
;; SEQ ID NO: 3455  
;; LENGTH: 1033  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-258-899A-3455

Query Match 21.1%; Score 187.5; DB 6; Length 1033;  
Best Local Similarity 31.9%; Pred. No. 0.075; Indels 23; Gaps 7;

Matches 60; Conservative 29; Mismatches 76; Indels 23; Gaps 7;

QY 5 REEAEEAKRRADAKEOGKPRGAKGVPGELATPPDKENDAKSSDSSVGEETLPPSPSLK 64  
DB 712 KEBAKSPKAKSPVKEBAKSPKAKSPVKEAKTPPEKAKSPVKEBAKS-----PEKAKS 765  
QY 65 PEKVAEAKKVEBAKKAEDQKEEDRNYPNTYKLELEI-----AESDV--EVKAE 117  
DB 766 PE-KAKTIDVKSPEAKTPAKEARSPADKPEKAKSPVKEEVKSPKAKSPDKEDAKAPE 824  
QY 118 LEL-VKEAKPRNE-----KYQAKAEVSKAEATRLKIKTDRAKAEBAKRAA 170  
DB 825 KEIPKKEVKSPVKEBKQEVKVEPKKAEBAKAPATPTEBKSKK--EBAKPKKA 882  
QY 171 EEDVKER 178  
DB 883 PKPKVEK 890

RESULT 5  
US-10-425-114A-47285

;; Sequence 47285, Application US/10425114A  
;; GENERAL INFORMATION:

;; APPLICANT: Liu, Jingdong  
;; APPLICANT: Zhou, Yihua  
;; APPLICANT: Kovalic, David K.  
;; APPLICANT: Screen, Steven E  
;; APPLICANT: Tabaska, Jack E  
;; APPLICANT: Cao, Yongwei  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
;; FILE REFERENCE: 38-21(53113)B  
;; CURRENT APPLICATION NUMBER: US/10/425,114A  
;; CURRENT FILING DATE: 2003-04-28  
;; NUMBER OF SEQ ID NOS: 73128  
;; SEQ ID NO: 47285  
;; LENGTH: 507  
;; TYPE: PRT  
;; ORGANISM: Glycine max  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: 700863165\_F11.pcp  
US-10-425-114A-47285

Query Match 18.3%; Score 162.5; DB 6; Length 507;  
Best Local Similarity 28.2%; Pred. No. 0.56; Indels 25; Gaps 6;

Matches 55; Conservative 39; Mismatches 76; Indels 25; Gaps 6;

QY 2 KTDREAEBAKRRADAKEOGKPRGAKGVPGELATPPDKENDAKSSDSSVGE-----56  
DB 262 KEERKEEBAKRRQOEERKEKRGEEBAKREBEBAKME--KERERKEEBAEREEBAAR 320  
QY 57 -----TLPSLSLPEKKVAEAKKVEBAKKAEDQKEEDRNYPNTYKLELEIASED 110  
DB 321 EOEWRKKEBAAROEOKERRROEBEBAKMAEKERQEO-----ATREAEQORE 374  
QY 111 VEVKAEEL-----ELVKEBAKPRNEBKVKQAKAEVE--SKAEATRLKIKTDRAKAE-E 163

DB 375 QOAKREAAQREGBAKREGRRRREEVAGKGEEHEHAGHBRKKEEFTARDEQOE 434  
QY 164 EAKKAAEEDVKER 178  
DB 435 QAERQOQKREKRROK 449

RESULT 6  
PCT-US03-11802-20

;; Sequence 20, Application PC/TUS0311802  
;; GENERAL INFORMATION:

;; APPLICANT: Washington University  
;; TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protectiv  
;; TITLE OF INVENTION: Immunogenicity  
;; FILE REFERENCE: 56029-40434  
;; CURRENT APPLICATION NUMBER: PCT/US03/11802  
;; PRIOR FILING DATE: 2003-04-15  
;; PRIOR APPLICATION NUMBER: US 60/373,626  
;; PRIOR FILING DATE: 2002-04-18  
;; PRIOR APPLICATION NUMBER: US 60/372,616  
;; NUMBER OF SEQ ID NOS: 21  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 20  
;; LENGTH: 257  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
PCT-US03-11802-20

Query Match 17.8%; Score 158.5; DB 1; Length 257;  
Best Local Similarity 29.5%; Pred. No. 0.55; Indels 41; Gaps 10;

Matches 56; Conservative 43; Mismatches 50; Indels 41; Gaps 10;

QY 2 KTDREAE-----EBAKRRADAKEOGKPRGAKGV-----PELATPPDKENDAKSSDSSVG 54  
DB 87 KAAKDADKMIIDEKKR-----EEBAKTKFNTVRAMVPEPEQLAETKKKSEBAKQK-----138  
QY 55 EETLPSLSLPEKKVAEAKKVEBAKKAEDQKEE--DRRNYPNTYKLELEIASEDEV 113  
DB 139 -----APEL--TKLEBAKAKLEBAEKATBAKQKDAEAVAP-----QAKIAELENOV 185  
QY 114 KKALELEVKEBAKPRNEBKVKOA-----KAIVESKAEATRLKIKTDRAKAE 163  
DB 186 HRLBOEL--KEIDSESEDYAKGFRAPLOSKDAAKAKSKLELSDKIDELDAETAKL 243  
QY 164 EAKKAAEED 173  
DB 244 EDLKAKEEN 253

RESULT 7  
PCT-US03-11802-21

;; Sequence 21, Application PC/TUS0311802  
;; GENERAL INFORMATION:

;; APPLICANT: Washington University  
;; TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protectiv  
;; TITLE OF INVENTION: Immunogenicity  
;; FILE REFERENCE: 56029-40434  
;; CURRENT APPLICATION NUMBER: PCT/US03/11802  
;; PRIOR FILING DATE: 2003-04-15  
;; PRIOR APPLICATION NUMBER: US 60/373,626  
;; PRIOR FILING DATE: 2002-04-18  
;; PRIOR APPLICATION NUMBER: US 60/372,616  
;; PRIOR FILING DATE: 2002-04-15  
;; NUMBER OF SEQ ID NOS: 21  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 21  
;; LENGTH: 467  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
PCT-US03-11802-21



```

: CURRENT FILING DATE: 2003-11-06
: PRIOR APPLICATION NUMBER: PCT/US01/04098
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: 09/774,434
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: 09/728,422
: PRIOR FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: 09/693,325
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 09/663,561
: PRIOR FILING DATE: 2000-09-15
: PRIOR APPLICATION NUMBER: 09/654,936
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 09/620,325
: PRIOR FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: 09/598,075
: PRIOR FILING DATE: 2000-06-20
: PRIOR APPLICATION NUMBER: 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 09/496,914
: PRIOR FILING DATE: 2000-02-03
: NUMBER OF SEQ ID NOS: 3960
: SOFTWARE: Custom
: SEQ ID NO: 1881
: LENGTH: 725
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-258-899A-1881

Query Match 16.6%; Score 148; DB 6; Length 725;
Best Local Similarity 25.5%; Pred. No. 3;
Matches 52; Conservative 47; Mismatches 63; Indels 42; Gaps 8

QY 1 IKTRDEAEAEAKRRADAKEQGRKGRAGVPEGLATPD---KKENDAKSSDSVVG- 54
Db 61 LNDIFEANSDDSQOSETKEPTSSKKKKKKLROREESPDLLKKKAKAGLTKDSKDL 120
QY 55 EETLPSP-SIKPEKKVAAEAKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDVE 112
Db 121 ESSLESIVFDLRTKRISEAKEBELKESKKPKKDEVK-----TKELK-----K 163
QY 113 VKKAELELVK---EAKPEPNESEKVAQAKAEVSESKA-----EATRLKIKITDR 158
Db 164 VKKEIEIDLTKTKTEDPEKNKTKTKKEKFVESQVSESSVLNDSPPEDDSGLHSDRRE 223
QY 159 KKAEEEAARKRAEEDKVK---EKP 179
Db 224 KQNTKSARERRAGQDMGLEHGEKPK 247

RESULT 13
US-10-258-899A-3849
: Sequence 3849, Application US/10258899A
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Drmanac, Radcoje T.
: APPLICANT: Asundi, Vinod
: APPLICANT: Zhou, Ping
: APPLICANT: Xu, Chongjun
: APPLICANT: Cao, Yicheng
: APPLICANT: Ma, Yungqing
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Duntui
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyan
: APPLICANT: Chen, Rui-hong
: APPLICANT: Wang, Zhi Wei
: APPLICANT: Xue, Aiding
: APPLICANT: Yang, Yonghong
: APPLICANT: Weinman, Tom
: APPLICANT: Goodrich, Ryle

```

```
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 787CIP2-26/US
/ CURRENT APPLICATION NUMBER: US/10/258,899A
/ CURRENT FILING DATE: 2003-11-06
/ PRIOR APPLICATION NUMBER: PCT/US01/04098
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 09/774,434
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 09/728,422
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: 09/693,325
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 09/663,561
/ PRIOR FILING DATE: 2000-09-15
/ PRIOR APPLICATION NUMBER: 09/654,936
/ PRIOR FILING DATE: 2000-09-01
/ PRIOR APPLICATION NUMBER: 09/620,325
/ PRIOR FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: 09/598,075
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/560,875
/ PRIOR FILING DATE: 2000-04-27
/ PRIOR APPLICATION NUMBER: 09/496,914
/ PRIOR FILING DATE: 2000-02-03
/ NUMBER OF SEQ ID NOS: 3960
/ SOFTWARE: Custom
/ SEQ ID NO: 3849
/ LENGTH: 774
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)...(774)
/ OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-258-899A-3849
```

```
Query Match 16.6%; Score 148; DB 6; Length 774;
Best Local Similarity 25.5%; Pred. No. 3.1;
Matches 52; Conservative 47; Mismatches 63; Indels 42; Gaps 8;
```

```
QY 1 KTDREAEAEARRADAEQGPGRAGVPGELATPD---KENDAKSDSSVG-- 54
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 152 LNDIFEAANSDDQSEFTKEDTSPKPKKKKLRQREKSPDDLKKKAKAGLKKKSPDL 211
QY 55 EETLPSR--SLKPEKKVAEAEKVEAEAKKADQKEEDRRNYPTNTYTLLELAESVVE 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 212 ESSLESLVFDLRTKRRISEAEELKESKKPKKDEVE-----TKELK-----K 254
QY 113 VKKAELELVK---EAEKPRNEEKVKQAKAEVSKA-----EATRLKIKITDR 158
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 255 VKKEIRLPLKTRTDPKRNKTKKEKFEVSESVLSDSPPEDDSGHSDSREE 314
QY 159 KKAEEAEAKKAEDDKV--EKP 179
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 KONTKSARERAGQDWGLEHGEKPK 338
```

```
RESULT 14
US-09-820-843B-95
/ Sequence 95, Application US/09820843B
/ GENERAL INFORMATION:
/ APPLICANT: Council of Scientific and Industrial Research
/ TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE
/ FILE REFERENCE: 063915
/ CURRENT APPLICATION NUMBER: US/09/820,843B
/ CURRENT FILING DATE: 2001-03-30
/ NUMBER OF SEQ ID NOS: 118
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO: 95
/ LENGTH: 369
/ TYPE: PRT
/ ORGANISM: T. pallidum
```

```
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: conserved hypothetical protein
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: g1|3322751
US-09-820-843B-95
```

```
Query Match 16.4%; Score 146; DB 5; Length 369;
Best Local Similarity 29.7%; Pred. No. 2.4;
Matches 52; Conservative 29; Mismatches 56; Indels 38; Gaps 7;
```

```
QY 5 REAEAEARRADAEQGPGRAGVPGELATPDKKENDAKSDSSVGSETLPSPLK 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 193 RKEAE--ARRKEAEARRKAEAEARRKE-----EARRKAEARRKAEARRKAE 246
QY 65 PEKVAAE--AEKVAEAKKAEADQKEEDRRNYPTNTYTLLELAESDVEYKAELELVK 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 247 ARKAEAEARRKAEAEARRK---EAEARR-----KEAE-EARR 281
QY 123 EAEKPRNEEKVKQAKAEVSKAATRLKIKITDRKAEAEAKKAEDDKVE 177
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 282 KEAEAEARRK---EAEARRKAEAEARRKAEAEARRKAEAEARRK 331
```

```
RESULT 15
US-09-820-843B-107
/ Sequence 107, Application US/09820843B
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Council of Scientific and Industrial Research
/ TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE
/ FILE REFERENCE: 063915
/ CURRENT APPLICATION NUMBER: US/09/820,843B
/ CURRENT FILING DATE: 2001-03-30
/ NUMBER OF SEQ ID NOS: 118
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO: 107
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Plasmodium falciparum
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: hypothetical protein
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: g1|3845248
US-09-820-843B-107
```

```
Query Match 16.4%; Score 146; DB 5; Length 665;
Best Local Similarity 24.4%; Pred. No. 3.4;
Matches 52; Conservative 39; Mismatches 82; Indels 40; Gaps 6;
```

```
QY 2 KTDREAEAEARRADAEQGPGRAGVPGELATPDKKENDAKSDSSVGSETLPSP 61
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 159 KKGQDINSNAENKQDYKGVKLEBEKK--BEKISDQKVEENKKSDDHKVEENKSD 215
QY 62 SLKPEKVAAEAEKVEAEAKK--KAEDQKEEDRR-----NYTNTYKTLLE--E 105
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 DHKVEENKKSDDHKIEEVKVEEHEDEEDKKEKSENKKNQDENEDNDEISDED 275
QY 106 IAEVDV-----VKKAELELVAEAEKPRNEEKVKQAKAEVSKK 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 276 EVDVDEEDKKNENDIDDDKKTETDKTLLEENENIIKEKFSDDKQGNKNDTKKE--KSKD 334
QY 146 AEATRLKIKITDRKAEAEAKKAEDDKVEK 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 335 TEKEKSKDIEKSKDKKEKSKDKKEKSKDK 367
```

```
Search completed: November 21, 2003, 13:41:39
Job time : 8.03329 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:23:06 ; Search time 5.20984 Seconds

(without alignments)  
1624.773 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442

Perfect score: 889  
Sequence: 1 IKTDREAEAEAEAKRADAKE.....AEEAEAKAEADPKVEKPA 180

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	188.5	21.2	1020	1	NFH_HUMAN
2	182.5	20.5	771	1	CALD_CHICK
3	181.5	20.4	1403	1	YDF3_SCHPO
4	177	19.9	1087	1	NFM_MOUSE
5	176.5	19.9	644	1	NFM_RABIT
6	176	19.8	1178	1	NMNA_YEAST
7	172.5	19.4	831	1	NFH_RAT
8	172	19.3	1220	1	IF2P_HUMAN
9	171.5	19.3	805	1	IF2P_AQUAE
10	170.5	19.2	344	1	MST1_DROHY
11	169.5	19.1	1391	1	MST2_DROHY
12	169	19.0	1359	1	ATRX_CAEEL
13	168.5	19.0	407	1	IE68_HSVSA
14	168.5	19.0	2468	1	MAPB_HUMAN
15	166.5	18.7	845	1	NFM_RAT
16	166.5	18.7	6632	1	UN65_CAEEL
17	163.5	18.4	2459	1	MAPB_RAT
18	160	18.0	667	1	CYL1_BOVIN
19	159.5	17.9	721	1	ENP1_TORCA
20	159	17.9	848	1	NFM_MOUSE
21	158	17.8	421	1	TOLA_ECOLI
22	158	17.8	1240	1	YNU1_YEAST
23	158	17.8	2464	1	MAPB_MOUSE
24	156	17.5	810	1	NFM_BOVIN
25	155.5	17.5	1002	1	IF2P_YEAST
26	154.5	17.4	877	1	INCE_CHICK
27	152.5	17.2	488	1	CYL2_BOVIN
28	152.5	17.2	919	1	INCE_HUMAN
29	152	17.1	793	1	CALD_HUMAN
30	151	17.0	705	1	TRDN_RABIT
31	150.5	16.9	857	1	NFM_CHICK
32	150	16.9	700	1	TRDN_CANFA
33	149.5	16.8	678	1	GARP_PLAFA

34	149.5	16.8	905	1	IF2_VIBPA	Q87m02 vibrio para
35	148.5	16.7	538	1	CYL1_HUMAN	P35663 homo sapien
36	146.5	16.5	880	1	INCE_MOUSE	Q9wue2 mus musculu
37	146	16.4	284	1	TPM1_CHICK	P04268 gallus galli
38	146	16.4	284	1	TPM1_COTJA	P58773 coturnix co
39	146	16.4	348	1	CYL2_HUMAN	Q14093 homo sapien
40	146	16.4	938	1	CAPA_HUMAN	Q13111 homo sapien
41	145.5	16.4	414	1	NSR1_YEAST	P27476 saccharomyc
42	145.5	16.4	2492	1	ATRX_HUMAN	P46100 homo sapien
43	145	16.3	325	1	TMPP_TREPA	P19649 treponema p
44	145	16.3	2116	1	MYS2_DICDI	P08799 dictyosteli
45	143.5	16.1	4684	1	PLE1_HUMAN	Q15149 homo sapien

## ALIGNMENTS

RESULT 1  
NFH\_HUMAN  
ID NFH\_HUMAN STANDARD; PRT; 1020 AA.  
AC P12036;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurofilament triplet H protein (200 kDa neurofilament protein)  
DE (Neurofilament heavy polypeptide) (NF-H).  
GN NFH OR NFH.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8328981; Pubmed=3138108;  
RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;  
RT "The structure and organization of the human heavy neurofilament  
RT subunit (NF-H) and the gene encoding it.";  
RL EMBO J. 7:1947-1955 (1988).  
CC CC  
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P. NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENTLY WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL, curation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X15306; CAA33366.1; -.  
CC EMBL; X15307; CAA33366.1; JOINED.  
CC EMBL; X15308; CAA33366.1; JOINED.  
CC EMBL; X15309; CAA33366.1; JOINED.  
CC PIR; S00979; OPHUH.  
CC GeneW; HGNC:7737; NEFH.  
CC MIM; 162230; -.  
CC GO; GO:0005883; C:neurofilament; NAS.  
CC GO; GO:0007399; P:neurogenesis; NAS.  
CC InterPro; IPR001664; IF.  
CC Pfam; PF00038; filament; 1.  
CC PROSITE; PS00226; IF; 1.

[illegible]

## RESULT 2

ID CAD CHICK STANDARD. PRT. 771 AA.  
AC P12957; 003698; Q90756; Q90761; Q92018; Q99230;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Caldesmon (CDM).  
GN CALDI OR CAD.  
OS Gallus gallus (Chicken).  
OC Archaeopta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Eukaryota; Aves; Neognathae; Galliformes; Phasianidae; Phasianini  
OC Gallus.  
CX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).  
RC TISSUE=Gizzard; and Oviduct;  
RX MEDLINE=89340480; PubMed=2760048;  
RA Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.,  
RT "Cloning and expression of a smooth muscle caldesmon.";  
RL J. Biol. Chem. 264:13873-13879(1989).  
RN [2]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).  
RC TISSUE=Gizzard;  
RX MEDLINE=90026426; PubMed=2803315;  
RA Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;  
RT "Primary structure and functional expression of h-caldesmon  
complementary DNA.";  
RL Biochem. Biophys. Res. Commun. 164:503-511(1989).  
RN [3]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).  
RC TISSUE=Brain;  
RX MEDLINE=91093148; PubMed=1824698;  
RA Hayashi K., Fujio Y., Kato I., Sobue K.;  
RT "Structural and functional relationships between h- and  
l-caldesmons.";  
RL J. Biol. Chem. 266:355-361(1991).  
RN [4]  
RP SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD)

CC MEDLINE=94071934; PubMed=82505919;  
RA Hatuna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;  
RT "Common structural and expressional properties of vertebrate  
RL caldesmon genes.";  
RN Biochem. Biophys. Res. Commun. 197.145-153(1993).  
RP [5]  
RC SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).  
RX TISSUE=Gizzard;  
RY MEDLINE=92042686; PubMed=1939602;  
RZ Bryan J., Lee R.;  
RT "Sequence of an avian non-muscle caldesmon";  
RN J. Muscle Res. Cell Motil. 12:372-375(1991).  
RP [6]  
RC SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).  
RX TISSUE=Brain;  
RY MEDLINE=94271210; PubMed=8002994;  
RZ Yano H., Hayashi K., Hatuna M., Sobue K.;  
RT "Identification of two distinct promoters in the chicken caldesmon  
RL gene.";  
RN Biochem. Biophys. Res. Commun. 201:618-626(1994).  
RP [7]  
RC SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.  
RX TISSUE=Gizzard;  
RY MEDLINE=89273666; PubMed=2730665;  
RZ Hayashi K., Yamada S., Kanda K., Kimitsuka F., Kato I., Sobue K.;  
RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of  
RN the tropomyosin-binding domain in tropomyosin T.";  
RP Biochem. Biophys. Res. Commun. 161:38-45(1989).  
RL [8]  
RC SEQUENCE OF 498-525.  
RX MEDLINE=88293484; PubMed=3401222;  
RZ Mornet D., Audemard E., Derancourt J.;  
RT "Identification of a 15 kilodalton actin binding region on gizzard  
RN caldesmon probed by chemical cross-linking".  
RP Biochem. Biophys. Res. Commun. 154:564-571(1988).  
RL [9]  
RC PHOSPHORYLATION SITES.  
RX MEDLINE=92041815; PubMed=1939059;  
RZ Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;  
RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of  
RN phosphorylation sites";  
RP J. Biol. Chem. 266:119971-119975(1991).  
RL [10]  
RC PHOSPHORYLATION OF TYR-27 AND TYR-165.  
RX MEDLINE=20026923; PubMed=10559216;  
RZ Wang Z., Danielson A.J., Mailhe N.J., McMann M.J.;  
RT "Tyrosine phosphorylation of caldesmon is required for binding to the  
RN Shc-Grd complex".  
RP J. Biol. Chem. 274:33807-33813(1999).  
RL -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE  
CC REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND  
CC NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN  
CC FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH  
CC INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCULAR  
CC TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.  
CC THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS  
CC POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO  
CC MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN  
CC ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPING.  
CC SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON  
CC STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative products; Named isoforms=3;  
CC Name=Gizzard h-cad;  
CC IsoId=P12957-1; Sequence=Displayed;  
CC Name=brain l-cad;  
CC IsoId=P12957-2; Sequence=VSP\_004152, VSP\_004153,  
CC Name=Gizzard l-cad;  
CC IsoId=P12957-3; Sequence=VSP\_004153;  
CC -1- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)  
CC IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-  
CC MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN  
CC NON-MUSCLE TISSUES AND CELLS NOT EXPRESSED IN SKELETAL MUSCLE OR



CC HEART.  
 CC -1- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-  
 CC BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-  
 CC BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL  
 CC HELICAL REGION IN THE MUSCLE FORMS.  
 CC -1- PTM: Phosphorylation in non-muscle cells. Phosphorylation by CDC2  
 CC during mitosis causes caldesmon to dissociate from microfilaments.  
 CC Phosphorylation reduces caldesmon binding to Actin, Myosin, and  
 CC Calmodulin as well as its inhibition of actomyosin ATPase  
 CC activity. Phosphorylation also occurs in both quiescent and  
 CC dividing smooth muscle cells with similar effects on the  
 CC interaction with Actin and Calmodulin and on microfilaments  
 CC reorganization (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE CALDESMON FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 CC EMBL; J04968; AAA49067.1; -;  
 CC EMBL; D17648; BAA04539.1; JOINED.  
 CC EMBL; D17634; BAA04539.1; JOINED.  
 CC EMBL; D17637; BAA04539.1; JOINED.  
 CC EMBL; D17637; BAA04539.1; JOINED.  
 CC EMBL; D17637; BAA04539.1; JOINED.  
 CC EMBL; D17639; BAA04539.1; JOINED.  
 CC EMBL; D17640; BAA04539.1; JOINED.  
 CC EMBL; D17641; BAA04539.1; JOINED.  
 CC EMBL; D17642; BAA04539.1; JOINED.  
 CC EMBL; D17643; BAA04539.1; JOINED.  
 CC EMBL; D17644; BAA04539.1; JOINED.  
 CC EMBL; D17645; BAA04539.1; JOINED.  
 CC EMBL; D17646; BAA04539.1; JOINED.  
 CC EMBL; M28417; AAA48810.1; -;  
 CC EMBL; M60620; AAA48936.1; -;  
 CC EMBL; D17648; BAA04538.1; -;  
 CC EMBL; D17635; BAA04538.1; JOINED.  
 CC EMBL; D17636; BAA04538.1; JOINED.  
 CC EMBL; D17637; BAA04538.1; JOINED.  
 CC EMBL; D17639; BAA04538.1; JOINED.  
 CC EMBL; D17640; BAA04538.1; JOINED.  
 CC EMBL; D17641; BAA04538.1; JOINED.  
 CC EMBL; D17642; BAA04538.1; JOINED.  
 CC EMBL; D17643; BAA04538.1; JOINED.  
 CC EMBL; D17644; BAA04538.1; JOINED.  
 CC EMBL; D17645; BAA04538.1; JOINED.  
 CC EMBL; D17646; BAA04538.1; JOINED.  
 CC EMBL; D17647; BAA04538.1; JOINED.  
 CC EMBL; D17648; BAA04540.1; JOINED.  
 CC EMBL; D17634; BAA04540.1; JOINED.  
 CC EMBL; D17636; BAA04540.1; JOINED.  
 CC EMBL; D17637; BAA04540.1; JOINED.  
 CC EMBL; D17639; BAA04540.1; JOINED.  
 CC EMBL; D17640; BAA04540.1; JOINED.  
 CC EMBL; D17641; BAA04540.1; JOINED.  
 CC EMBL; D17642; BAA04540.1; JOINED.  
 CC EMBL; D17643; BAA04540.1; JOINED.  
 CC EMBL; D17644; BAA04540.1; JOINED.  
 CC EMBL; D17645; BAA04540.1; JOINED.  
 CC EMBL; D17646; BAA04540.1; JOINED.  
 CC EMBL; D17647; BAA04540.1; JOINED.  
 CC EMBL; D17648; BAA04540.1; JOINED.  
 CC EMBL; M59762; AAA48649.1; -;  
 CC EMBL; D17552; BAA04490.1; -;  
 CC EMBL; M26684; AAA48811.1; -;  
 CC PIR; A33430; A33430.  
 CC HSP; P05412; IFOS.  
 CC InterPro; IPR006017; Caldesmon.  
 CC InterPro; IPR006018; Caldesmon\_LSP.  
 CC Pfam; PF02029; Caldesmon; 1.

DR PRINTS; PRO1076; CALDESMON.  
 KW Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;  
 RV Repeat; Alternative splicing.  
 FT DOMAIN 26 199  
 FT MYOSIN AND CALMODULIN-BINDING.  
 FT 10 X 13 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 251 390  
 FT REPEAT 251 265  
 FT REPEAT 266 278  
 FT REPEAT 279 291  
 FT REPEAT 294 306  
 FT REPEAT 309 321  
 FT REPEAT 324 336  
 FT REPEAT 337 349  
 FT REPEAT 350 362  
 FT REPEAT 363 375  
 FT REPEAT 378 390  
 FT REPEAT 523 580  
 FT DOMAIN 622 632  
 FT TROPOMYOSIN-BINDING (POTENTIAL).  
 FT TROPOMYOSIN-BINDING (POTENTIAL).  
 FT  
 Query Match 20.5%; Score 182.5; DB 1; Length 771;  
 Best Local Similarity 33.3%; Pred. No. 0.017; Indels 23; Gaps 6;  
 Matches 61; Conservative 28; Mismatches 71;  
 OY 5 REAAEEBAKRRADAKKQKPKRAKRGVPGELATPD-----KKENDAKSSDSSVGERT 57  
 DB 171 RQDEEEGCKKEKXSEEEKPK-----EVPTEBNQVDVAVEKSTDKKEVEVETKTLAVNEN 225  
 OY 58 LPSPSLAPEKKVAE-AEKVVEAKKAKEDQKEPRNYPPTTYTLEFIASDVEYKA 116  
 DB 226 DTNMLRGEOSITPAADKE 285  
 OY 117 ELELVKEBAK---BPRNEKYKQAKAEVESKKAETRLKTKTRKKAEEBAK---RKA 169  
 DB 286 E---ERRAKAEEKRAAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKA 342  
 OY 170 AEE 172  
 DB 343 AEE 345  
 RESULT 3  
 ID YDF3 SCHPO STANDARD; PRT; 1403 AA.  
 AC 010475;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable eukaryotic initiation factor C17C9.03.  
 GN SPAC17C9.03.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 CX NCBI\_TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Scouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,  
 RA Weltegens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Filicz C., Holzer E., Moesti D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,

```

RA Eger P, Elzemerhmann W., Medler H., Mambutt R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado I., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesery D., Barrett B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
-1- SIMILARITY: TO YEAST EUKARYOTIC INITIATION FACTOR 4F SUBUNIT'S P130
AND P150.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73099; CAA97349.1; -.
DR PIR; T11583; T11583.
DR GeneDB, SPombe; SPAC1C9.03; -.
DR InterPro; IPR003890; IF_eIF4G.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
KM Hypothetical protein; Initiation factor; Protein biosynthesis;
KM RNA-binding.
SQ SEQUENCE 1403 AA; 154034 MW; 0317EE65BE2A1B63 CRC64;

Query Match 20.4%; Score 181.5; DB 1; Length 1403;
Best Local Similarity 31.3%; Pred. No. 0.03;
Matches 56; Conservative 35; Mismatches 57; Indels 31; Gaps 7;

OY 5 REEAREEKKRRADAKGOGKPKGRARGVPGELATDPDKKENDKSSDSVGCEETLPSPLK 64
DB 562 KREAEQAKREAEAEAKREAEAEAKREAE--EKAKREAEENAKREAEAKRA----- 613
OY 65 PEKKVAEAEKVE--EAKKKAEDQKEDRRNVPYNTYKTLLEIAESDVEVKAELVKE 123
DB 614 EEKAKREAEAEKAKREAEAEKAKREAEAKR-----EAEKAKREAEAKR 659
OY 124 EAKEPRNEEKKYQAKAEV--ESKKAATRLKTKTD-----RKKAEEAKRKKAEDKYK 176
DB 660 EAENAKREAEKAKREAEENAKREAE--EKVKRETEENAKKKAEEGKREADKNPEIK 716

RESULT 4
NFH_MOUSE STANDARD; PRT; 1087 AA.
ID_NFH_MOUSE
AC P19246; O61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1982 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurofilament tripllet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
DE NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RX MEDLINE=69121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit.",
RL Gene 68:307-314(1988).
RN (2)
RX SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=1145094;

```

RA	Shneiderman P.S., Carden M.J., Lee J.F., Lazzerini R.A.;	19.98;	Score 177,	DB 1,	Length 1087,
RT	"The structure of the largest murine neurofilament protein (NF-H) as	33.7%;	Pred. No. 0.043;		
RI	revealed by cDNA and genomic sequences.";				
RL	Brain Res. 464:217-231(1988).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-Swiss Webster; TISSUE=Brain;				
RA	Carden M.J.;				
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.				
CC	-I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,				
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.				
CC	NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT				
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.				
CC	-I- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS				
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS				
CC	THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF				
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE				
CC	OF AXONAL CALIBER.				
CC	-I- PM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING				
CC	OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE				
CC	LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND				
CC	COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.				
CC	-I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.				
CC	-I- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534				
CC	TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; M24496; AAA38813.1; -				
DR	EMBL; M23449; AAA38813.1; JOINED.				
DR	EMBL; M24494; AAA38813.1; JOINED.				
DR	EMBL; M24495; AAA38813.1; JOINED.				
DR	EMBL; M35131; AAA39809.1; ALT_FRAME.				
DR	EMBL; Z31012; CAA83229.1; -				
DR	PIR; J70368; QPM5H.				
DR	MGI: 97309; Neth.				
DR	InterPro: IPR001664; IF.				
DR	Pfam: PF00038; Filament; 1.				
DR	PROSITE; PS00226; IF, 1.				
KW	Intermediate filament; Coiled coil; Neutrone; Phosphorylation;				
KW	Repeat.				
FT	DOMAIN	1	97	HEAD.	
FT	DOMAIN	98	408	ROD.	
FT	DOMAIN	409	1087	TAIL.	
FT	DOMAIN	436	517	GLU-RICH (ACIDIC).	
FT	DOMAIN	519	886	50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.	
FT	DOMAIN	887	1087	GLU/LYS-RICH.	
FT	DOMAIN	98	129	COIL 1A.	
FT	DOMAIN	130	141	LINKER 1.	
FT	DOMAIN	142	239	COIL 1B.	
FT	DOMAIN	240	261	LINKER 12.	
FT	DOMAIN	262	283	COIL 2A.	
FT	DOMAIN	284	287	LINKER 2.	
FT	DOMAIN	288	408	COIL 2B.	
FT	CONFLICT	133	133	K -> QA (IN REF. 2 AND 3).	
FT	CONFLICT	199	199	A -> AR (IN REF. 2 AND 3).	
FT	CONFLICT	281	281	S -> T (IN REF. 2 AND 3).	
FT	CONFLICT	492	492	L -> G (IN REF. 2 AND 3).	
FT	CONFLICT	551	551	P -> BREAKSP (IN REF. 3).	
FT	CONFLICT	689	712	MISSING (IN REF. 3).	
FT	CONFLICT	714	714	G -> A (IN REF. 3).	
FT	CONFLICT	814	814	V -> M (IN REF. 2 AND 3).	
FT	CONFLICT	843	843	T -> N (IN REF. 2 AND 3).	
SO	SEQUENCE	1087 AA;	116612 MW;	57BAC76A38ED1CB9 CRC64;	

Matches	65;	Conservative	21;	Mismatches	73;	Indels	34;	Gaps	9;
Qy	5	REEAEFEKRRADAKEOGQKPKGRAR	GYPGELATDDKKENDAKSSDSSVGEE	-TLPSPSL	63				
Db	859	KEKASPEKEAEKTEKAEVAPKDE	-----EVSFVKEEVAKAEPPKVEEKTLP	PTP--	909				
Qy	64	KPEKVAEAEKKEVSAKKKAEDQCKEDR	RNYPNTNYKTLLELAESDVYVK	-----ABLE	119				
Db	910	KTEAK----	ESKKDAPPAEPKPKVEEKTPTTEKPKOSTLEAKKEGKKKAVASBEE	965					
Qy	120	L-----VKEAKEPNEEKVKCAKAEVSKA	-----EATRLKIKITDPRKKAABEAK	-----	166				
Db	966	TPAKLVGVEEAKP----	KEKTEPTTKTEADDTAKESKPTTEKPKKEMPAPAEKDKTKE	10222					
Qy	167	RKAABEDYVKEKP	179						
Db	1023	EKTESRKEPEKP	1035						
RESULT 5									
AC	NEW_RABBIT	STANDARD;	PRT;	644	AA.				
AC	P54538;								
DT	01-OCT-1996	(Rel. 34, Created)							
DT	01-OCT-1996	(Rel. 34, Last sequence update)							
DT	16-OCT-2001	(Rel. 40, Last annotation update)							
DE	Neurofilament triplet M protein (160 kDa neurofilament protein)								
DE	(Neurofilament medium polypeptide) (NF-M) (Fragment).								
DN	NEF3 OR NEFM OR NFM.								
OS	Oryctolagus cuniculus (Rabbit).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Euteria; Lagomorpha; Leporidae; Oryctolagus.								
ON	NCBI_TaxID=9986;								
RP	SEQUENCE FROM N.A.								
RC	TISSUE=Heart;								
RX	MEDLINE=97055255; Pubmed=8899542;								
RA	Vladello M., Vettore S., Lamar E., Chien K.R., Gorza L.;								
RT	"Neurofilament M mRNA is expressed in conduction system myocytes of								
RT	the developing and adult rabbit heart."								
RL	J. Mol. Cell. Cardiol. 28:1833-1844(1996).								
CC	-1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,								
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.								
CC	-1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEW IS								
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS								
CC	THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF								
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE								
CC	OF AXONAL CALIBER.								
CC	-1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING								
CC	OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE								
CC	LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND								
CC	COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.								
CC	-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.								
CC	-----								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See <a href="http://www.1sb-sib.ch/announce/">http://www.1sb-sib.ch/announce/</a>								
CC	or send an email to <a href="mailto:license@1sb-sib.ch">license@1sb-sib.ch</a> ).								
CC	-----								
DR	EMBL; 247378; CA87454.1; -								
DR	PIR; S55395; S55395.								
DR	InterPro; IPR001664; IF.								
DR	Pfam; PF00038; filament; 1.								
DR	PROSITE; PS00226; IF; 1.								
KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation;								
KW	Glycoprotein.								
FT	NON TER	1							
FT	DOMAIN	<1	197						
FT	DOMAIN	198	643						
FT	DOMAIN	<1	33						

```

FT DOMAIN 34 50 LINKER 12.
FT DOMAIN 51 72 COIL 2A.
FT DOMAIN 73 76 LINKER 2.
FT DOMAIN 77 197 COIL 2B.
FT CARBOHD 217 217 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 644 AA; 72450 MW; 030FDA6A22889678 CRC64;

Query Match 19.9%; Score 176.5; DB 1; Length 644;
Best Local Similarity 29.3%; Fred. No. 0.029;
Matches 63; Conservative 30; Mismatches 85; Indels 37; Gaps 7

QY 1 IKTRDEEEAEAKRADAEQGGKPKGRKRVGPEGLAPPD--KKENDAKSSDSVGEETLP 59
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 268 VSVVEEEKEEAEKEEGEQAEEVAAAKKS--PVKATTPPEIKEEGEGEAEQEEEEE 326
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 SPSLR-----PEKKVAEAEKKVVEA-----KKKADOKEE--DRNY 94
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 327 DEGVKSDQAEBCGSEKESGSKNEGEQEGERTAEAGEVEAEAEKEKTEKSEVAAKE 386
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 95 PNTNYKTLEELASDVEVKAELVVEAEKEPRNEKKVQAKAEVSKA--EATRL 152
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 387 PVTEAKGKPKPKAKSPVPKSPVEEVKPVAEATAGCGEKEEKEVEKKVAAKSPDE 446
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 153 KIK-----TDKKAEAEAKRKRAAEEDKTKKXP 179
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 447 KVEKKEKPKDVPKKKASPVKEAEAEAAITTPK 481
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
MNNA_YEAST STANDARD; PRI: 1178 AA.
ID_MNNA_YEAST STANDARD; PRI: 1178 AA.
AC P36044; P36043; P89095;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE MNNA protein.
GN MNNA OR YKL200C/YKL201C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxId:4932;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=91175967; PubMed=9023541;
RA Odani T., Shizuma Y.-I., Tanaka A., Jigami Y.;
RT "Cloning and analysis of the MNNA gene required for phosphorylation
  of N-linked oligosaccharides in Saccharomyces cerevisiae.";
RL Glycobiology 6:805-810(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
CC MANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- SIMILARITY: TO YEAST YJR061W.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.sib.ch).
CC -----
CC EMBL; D83006; BA01676.1; -
CC EMBL; D828201; CA082046.1; -

```

```

DR EMBL, Z28200; CAA82044.1; -
DR PIR, S78475; S78475.
DR SGD, S0001684; MNM4.
DR GO, GO:0006487; P-N-linked glycosylation; IMP.
DR GO, GO:0006493; P/O-linked glycosylation; IMP.
DR GO, GO:0006950; P:response to stress; IDA.
DR Transmembrane; Signal-anchor; Repeat.
KW DOMAIN
FT TRANSMEM 1 27
FT 28 48
FT 49 1178
FT 1022 1174
FT 1042 1174
FT 1042 1049
FT 1050 1057
FT 1058 1065
FT 1066 1073
FT 1074 1081
FT 1082 1089
FT 1090 1097
FT 1098 1105
FT 1106 1113
FT 1112 1121
FT 1122 1129
FT 1130 1137
FT 1138 1144
FT 1145 1152
FT 1153 1160
FT 1161 1168
FT 1169 1174
FT 1174 40
FT 1042 1045
FT 1050 1053
FT 1054 1057
FT 1058 1061
FT 1062 1065
FT 1066 1069
FT 1070 1073
FT 1074 1077
FT 1078 1081
FT 1082 1085
FT 1086 1089
FT 1094 1097
FT 1098 1101
FT 1102 1105
FT 1134 1137
FT 1157 1160
FT 1165 1168
FT 1178 AA; 139380 MM; BC0SDAE0AFCB282 CRC64;
SQ
Query Match 19.8%; Score 176; DB 1; Length 1178;
Best Local Similarity 26.4%; Pred. No. 0.051;
Matches 47; Conservative 39; Mismatches 48; Indels 44; Gaps 4;
QY 2 KTDREAEAEKRRADAKGQKPKRARGVPGELATDDKKENDAKSSDSSVGESTLSP 61
DB 1039 RREKKKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKK 1074
QY 62 SLRDEKVAEAKKVEAEAKKAADQKEDRRNYPTNTYKTLLELAESDVEKKALELV 121
DB 1075 -----KKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKK 1115
QY 122 KEAAK-PRNEKVAQAAVEESKKAATRLKTKTDKKAAEEAKRRAAEDKVKKEK 178
DB 1116 KNEDEENKKNDEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEEKKKEE 1173

```

```

DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
GN NEFH OR NEFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
EN
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89065087; PubMed=3143606;
RA Breen K.C., Robinson P.A., Wilson D., Anderson B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites.";
RL FEBS Lett. 241:213-218(1988).
[2]
RP SEQUENCE OF 37-831 FROM N.A.
RX MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT in situ detection.";
RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
[3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RX MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wilson D., Anderson B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H).";
RL FEBS Lett. 209:203-205(1986).
[4]
RP SEQUENCE OF 318-831 FROM N.A.
RX MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
-1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
SUBSERVED BY THE TWO SMALLER NF PROTEINS.
-1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NEFH IS
PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
THOUGHT THAT PHOSPHORYLATION OF NEFH RESULTS IN THE FORMATION OF
INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
OF AXONAL CALIBER.
-1- FUNCTION: NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M37227; AAA41693.1; ALT_FRAME.
DR EMBL, X13804; CAA32038.1; ALT_FRAME.
DR EMBL, M21964; AAA41695.1; -.
DR EMBL, J04517; AAA41692.1; -.
DR InterPro, IPR001664; IF.

```

DR PFam: PF00038; filament, 1.  
 DR PROSITE: PS00226; IFP; 1.  
 KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation; Repeat.

```

FT FT      1  

FT DOMAIN   1  

FT CONFLICT 276 641 51 X 3 AA TANDEN REPEATS OF K-S-P.  

FT CONFLICT 164 164 L -> I (IN REF. 2).  

FT CONFLICT 185 185 I -> S (IN REF. 2).  

FT CONFLICT 193 193 L -> T (IN REF. 2).  

FT CONFLICT 199 199 M -> N (IN REF. 2).  

FT CONFLICT 346 346 K -> T (IN REF. 1).  

FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).  

FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  

FT CONFLICT 485 485 P -> S (IN REF. 2).  

FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  

FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  

FT CONFLICT 727 727 A -> V (IN REF. 4).  

FT CONFLICT 757 759 AAP -> GST (IN REF. 4).  

FT CONFLICT 769 769 T -> L (IN REF. 2).  

FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).  

SQ SEQUENCE 831 AA; 89486 MM; 1B0973C3F1E7F68 CRC64;

Query Match          19.4%; Score 172.5; DB 1; Length 831;  

Best Local Similarity 29.1%; Pred. No. 0.056;  

Matches 60; Conservative 35; Mismatches 68; Indels 43; Gaps 7
```

OY	11	EAKRRADKEQCKPEKRAKGVGELAPDPKKEN-----DAKSSD---SSVGEEET	57
Db	494	EAKSPAEKSPAEAKSPAEKSPAEVSKSPAEKSGAKSLAEKSPAEKSPVKERI	553
OY	58	LPSLSLK-----PEKTVAEAERKV-----BEAKKKEDQKEEDRNRP	95
Db	554	KPPAEVNSPEPKASPMRKEAKSPPEAKTLDVKSPPAKPPAKEARRPADIRSEPVKSPA	613
OY	96	TNTTKTELELAESDVEVKKALELVKEAKEPRNEEKVKCAKAVESKCATLREKIX	155
Db	614	KEAKSPPEKE-----ETRTEKVMAPKKEEVNSPVEVAKKEPKVKEEKTATRTKTVK	667
OY	156	TDRK-KAEEBA-KRKAAEDVKVEKP	179
Db	668	ESKKDEAPKEAQKPAKEKEPLETERP	693

RESULT 8

ID	IF2P_HUMAN	STANDARD;	PRT;	1220 AA.
AC	060841; Q95805; Q9UR81; Q9UNW7;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	translation initiation factor IF-2.			
GN	IF2 OR KIAA0741.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain:			
RX	MEDLINE=99087487; PubMed=9872452;			
RA	Negase T., Ichikawa K.-I., Suyama M., Kikuno R., Miyajima N.,			
RA	Tanaka A., Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XI.			
RT	The complete sequences of 100 new cDNA clones from brain which code			
RT	for large proteins in vitro."			
RL	DNA Res. 5:277-286(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cervical carcinoma;			
RX	MEDLINE=99362399; PubMed=10432305;			
RA	Wilson S.A., Sieltro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,			
RA	Kotsopoulos E., Adanson C.S., Kingsman S.M., Kingeman A.J.,			
RT	Martin-Rendon E.;			
RT	"Cloning and characterization of hIF2, a human homologue of bacterial			

```
RT translation initiation factor 2 and its interaction with HIV-1
RL Biochem. J. 342:97-103(1999).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=99218282; PubMed=10200264;
RA Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Deyer T.E.;
RT "Universal conservation in translation initiation revealed by human
RT and archaeal homologs of bacterial translation initiation factor
RT IF2."
RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
RN [4]
RP SEQUENCE OF 89-1220 FROM N.A.
RC TISSUE=Testis;
RA Koshier K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) from the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 833-1220 FROM N.A.
RX MEDLINE=21064499; PubMed=11124703;
RA Stancliff F., Bertocco E., Toppo S., Diognardi R., Simonati B.,
RA Camarata N., Zimblett R., Lantirachi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RT yeast sequences."
RL Yeast 18:69-80(2001).
CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEKS TO
CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
CC EMBL; AB018284; BAA34461.1; -.
DR EMBL; AJ006776; CAB44357.1; -.
DR EMBL; AF078035; AAD16006.1; -.
DR EMBL; AL133563; CAB63717.1; -.
DR EMBL; AJ006412; CAA07018.1; -.
DR PIR; T43483; T43483.
DR MIM; 606086; -.
DR GO; GO:0003743; F:translation initiation factor activity; NAS.
GC; GO:0006446; P:regulation of translational initiation; NAS.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PR00009; GTP_EFTU_1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFACT.
DR TIGRFAMs; TIGR00231; small_gtp; 1.
DR PROSITE; PS01176; IF2; FALSE_NBS.
KW Initiation factor; Protein biosynthesis; GTP-binding.
NP BIND 638 645
FT DOMAIN 39 50 POLY-LYS.
FT DOMAIN 94 99 POLY-LYS.
FT DOMAIN 138 142 POLY-ASP.
FT DOMAIN 313 322 POLY-LYS.
FT DOMAIN 353 356 POLY-GLU.
FT DOMAIN 361 364 POLY-GLU.
FT DOMAIN 491 496 POLY-GLU.
FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 640 640 V-G: LOSS OF ACTIVITY IN VIVO. RETAINS FULL ACTIVITY IN VITRO.
FT MUTAGEN 706 706 H-2E: LOSS OF ACTIVITY, BOTH IN VIVO AND IN VITRO.
FT MUTAGEN 706 706 H-Q: LOSS OF ACTIVITY IN VIVO. PARTIAL ACTIVITY IN VITRO.
FT MUTAGEN 706 706 D-N: LOSS OF ACTIVITY, BOTH IN VIVO AND
```

```

DR InterPro: IPR000795; EF_GTPbind.
DR InterPro: IPR004161; EFTU_D2.
DR InterPro: IPR001178; IF2_N.
DR InterPro: IPR006847; IF2_N.
DR InterPro: IPR001806; Ras_cirsimfmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF03144; GTP_EFTU_D2; 2.
DR Pfam: PF04760; IF2_N; 2.
DR PRINTS: PR00315; ELONGINACT.
DR PRINTS: PR00449; RASTRNSFRMG.
DR ProDom: PD16100; IF2; 1.
DR TIGRFAMs: TIGR00487; IF-2; 1.
DR TIGRFAMs: TIGR00231; small_GTP; 1.
DR PROSITE: PS01176; IF2; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
KW DOMAIN.
FT NP_BIND 309 458 GTP (BY SIMILARITY).
FT NP_BIND 315 322 GTP (BY SIMILARITY).
FT NP_BIND 362 366 GTP (BY SIMILARITY).
FT NP_BIND 416 419 GTP (BY SIMILARITY).
SQ SEQUENCE 805 AA; 91600 MW; 8A1BF300C69CA51E CRC64;

Query Match 19.3%; Score 171.5; DB 1; Length 805;
Best Local Similarity 29.0%; Pred. No. 0.06; Mismatches 53; Indels 51; Gaps 8
Matches 56; Conservative 33;

QY 4 DREAEAEARRADAKEQKPKRAKRGVPELATPDKENDAKSDSSVGEETLPSPL 63
DB 61 EEKEKEVTEQQA-----PAEV--EKKEEKK--EEVIVEEVEEK-- 100
QY 64 KPEKTALEAKKVEEKKKAELQKEDRRNPNTNYKLF-----LEIASDVEYKAE 117
DB 101 KPEVITVEEIEKKKEEKEEKKPK-----KSVELLKEILEKKKEKEKKVE 149
QY 118 -----LELVKEAKEPENNEKRVQAKAEVSKKAELRLK-----IKTRDKAE 163
DB 150 KEKKEKRVVEYKKEKKEKKEKKEKKEKKEKPKIMSKKEEIMRKLHAEVKKKQK 209
QY 164 EARRKAAEDDKV 176
DB 210 REKEKKKEEYK 222

RESULT 10
MS11 DROXY STANDARD; PRT; 344 AA.
ID MS11_DROXY STANDARD; PRT; 344 AA.
AC 008695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein ms101(1).
GN MS101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC OCNCB1; Drosophilidae; Drosophila.
OX NCBI_TaxId=7224;
RN 11
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen J., Buemann H., Heinlein U.A.;
RT "The Drosophila hydei gene Dhmsc101(1) encodes a testis-specific,
RT repetitive, axoneme-associated protein with differential abundance in
RT Y chromosomal deletion mutant flies.";
RL Dev. Biol. 162:414-425(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC SPERMATID BUNDLES.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.

```





CC -----  
DR EMBL: AF134196; AAD55361.1; -  
DR EMBL: AF000196; AAC24256.1; -  
DR PIR: T34036; T34036.  
DR WormPep; B0041.7; CE17314.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICC; 1.  
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.  
FT NP\_BIND 496 503  
FT SITE 636 639  
FT DOMAIN 67 70  
FT DOMAIN 266 272  
FT DOMAIN 276 281  
FT DOMAIN 372 375  
FT DOMAIN 603 608  
FT DOMAIN 859 862  
FT CONFLICT 479 479 C->F (IN REF. 2).  
SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;  
Query Match 19.0%; Score 169; DB 1; Length 1359;  
Best Local Similarity 26.9%; Pred. No. 0.13;  
Matches 54; Conservative 40; Mismatches 81; Indels 26; Gaps 5;  
OY 2 KTDREAEERARRADAKEQCKPKGRKGVPGELATPDCKENDAKSSDSVGEETLPSPLK 61  
DB 80 KKSRRKAKSESESDSEDEEDKSKSKKVKQVQCKKSKKRTSSSEDESDSEDEDEEERQK 139  
OY 62 SLKPEKTV-----AAEKVVERA-----KKKAEDQCKEDRRNYPNTTYTTL 102  
DB 140 SKKSKSKTKTKOTSSSSSESESEERKVKSKKCKEKSVKRAETSESDSEDEDEKPSKSKG 199  
OY 103 ELEIASDVVEYKAELELVKEAKEPRNEKVKQAKAEVSEKKAATLEKIKTKTRKKA 162  
DB 200 LKTKAKSESE---SSEDEKEVKSKSKSKKVKVKKSESEDEAPKPKTKTERKSKTSSE 256  
OY 163 EEAK-RKAEDDKVE---KP 179  
DB 257 ESSESEKSDSEDEEKESSPKP 277  
RESULT 13  
1E68\_HSVSA STANDARD; PRT; 407 AA.  
ID 1E68\_HSVSA AC Q01042;  
AC 001042;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Immediate-early protein.  
GN 73 OR ECLF1.  
OS Herpesvirus saimiri (strain 11).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OC NCBI\_TaxID=10383;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9233688; PubMed=1321287;  
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
RA Honess R.W.;  
RT "Primary structure of the herpesvirus saimiri genome.";  
RT J. Virol. 66:5047-5058 (1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
RT organization between HVS and Epstein-Barr virus.";

RL Virology 188:296-310 (1992).  
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND  
CC HSV-2 (IE-68 (US1), EHv-1 65, EHv-4 (OR4), PRV RSP40, AND VZV 63.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: X64346; CAA45696.1; -  
DR EMBL: M86409; AAA46149.1; -  
DR EMBL: S76368; AAB21116.1; -  
KW Early protein.  
FT DOMAIN 60 241  
FT SEQUENCE 407 AA; 46617 MW; FPD399CA82CE136C CRC64;  
Query Match 19.0%; Score 168.5; DB 1; Length 407;  
Best Local Similarity 26.6%; Pred. No. 0.046;  
Matches 46; Conservative 44; Mismatches 50; Indels 33; Gaps 5;  
OY 5 REAEAEERARRADAKEQCKPKGRKGVPGELATPDCKENDAKSSDSVGEETLPSPLK 64  
DB 54 QQAALTEQRRREVEVEEGEE--RERG-----EEERGEGEGEGEGE-----E 95  
OY 65 PEKVAAEKKEVBEAKKKAEDQCKEDRRNYPNTTYTTLLEIASDVVEYKAELELYKEE 124  
DB 96 AEEBEAEKEAEKEAEAEAEAEAEAE-----EAEBAEAEAEAEAEAE 141  
OY 125 AKEPRNEKVKQAKAEVSEKKAATRLKIKTKDKKAEERKRAAEEDKVE 177  
DB 142 EAEEAEAEAEAEAEAEAEAEAEAEAEAE--EAEEAEAEAEAEAEAE 193  
RESULT 14  
MAPB\_HUMAN  
ID MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
AC P46821;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain  
DE LC1].  
GN MAP1B.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=95104835; PubMed=7806212;  
RA Lien L.-L., Feener C., Fischbach N., Kunkel L.M.;  
RT "Cloning of human microtubule-associated protein 1B and the  
RT identification of a related gene on chromosome 15.";  
RT Genomics 22:273-280 (1994).  
RL Genomics 22:273-280 (1994).  
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
CC TWO TUBULIN SUBUNITS IN THE POLYMER. AND THIS BRIDGING OF SUBUNITS  
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
CC STABILIZING MICROTUBULES.  
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
CC WITH MAP1A AND MAP1B PROTEINS.  
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
CC responsible for the binding of MAP1B to microtubules.  
CC -1- PWM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
CC from MAP1B by proteolytic processing. It is free to associate with  
CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
CC of MAP1B (By similarity).



```

CC -I SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06237; AAA18904.1; -.
DR Genew; HGNC:6836; MAP1B.
DR MIM; 157129; -.
DR GO; GO:0005875; C:microtubule associated complex; TAS.
DR InterPro; IPR000102; MAP1B_neuraxin.
DR Pfam; PF00414; MAP1B_neuraxin; 10.
DR PROSITE; PS00230; MAP1B_NEURAXIN; 6.
DR Microtubules; Repeat; Phosphorylation.
KW CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1878 1894 MAP1B 1.
FT REPEAT 1895 1911 MAP1B 2.
FT REPEAT 1912 1928 MAP1B 3.
FT REPEAT 1929 1945 MAP1B 4.
FT REPEAT 1946 1962 MAP1B 5.
FT REPEAT 1963 1979 MAP1B 6.
FT REPEAT 1997 2013 MAP1B 7.
FT REPEAT 2014 2030 MAP1B 8.
FT REPEAT 2031 2047 MAP1B 9.
FT REPEAT 2048 2064 MAP1B 10.
FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT KEE AND KEEI/V REPEATS).
SQ SEQUENCE 2468 AA; 270618 MW; 540839C8DF09D61 CRC64;
Query Match 19.0%; Score 168.5; DB 1; Length 2468;
Best Local Similarity 29.3%; Pred. No. 0.22;
Matches 63; Conservative 31; Mismatches 76; Indels 45; Gaps 7;
QY 2 KTDREAEAEKRRADAKGQK-PKGRKRGVPGELATPDCKENDAKSDSSVGEETLPS 60
DB 555 KSVKESKEKEPTETKVAHVHVKPKVESKEKVMKKDKPVETKTPSTKEKVSKEEPS 614
QY 61 P-----SLKPEKVAEAKKVEAKKKAEDQKEEDRNYPNTTY 99
DB 615 PVKAEVAKQATDVKPKAKKKEKTKVKEDKKEKPKKVEAKKKEK-----TPIK 670.
QY 100 KTLLELAESVQVYKAELELVKEBAKEPRNE-----EKYQAAVESSKAE 147
DB 671 KEKPKKEEYKKEVKK--EIKKEEKEPKKEKVEKKEPPKPEKKEVKEKKEKKEKE 727
QY 148 ATR-LEKIKTDKKAEE--EEAKKKAEDKVKER 178
DB 728 PKKEIKKLPKDAKKSSTPLSEAKKPAALKPVKPK 762

```

```

RT the middle molecular weight neurofilament protein.";
RL J. Neurosci. 7:2590-2599(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=92332596; PubMed=1321159;
RA Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT "Schwann cells of the myelin-forming phenotype express neurofilament
RL protein NF-M.";
RL J. Cell Biol. 118:397-410(1992).
RN [3]
RP PHOSPHORYLATION SITES, AND REVISION TO 500.
RX MEDLINE=92165797; PubMed=1537832;
RA Xu Z.-S., Liu W.-S., Willard M.B.;
RT "Identification of six phosphorylation sites in the COOH-terminal
RL tail region of the rat neurofilament protein M.";
RL J. Biol. Chem. 267:4467-4471(1992).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=93346421; PubMed=8344946;
RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA Hart G.W.;
RT "Glycosylation of mammalian neurofilaments. Localization of multiple
RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RL L and M.";
RL J. Biol. Chem. 268:16679-16687(1993).
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18628; AAA41696.1; -.
DR EMBL; Z12152; CAA78136.1; -.
DR PIR; A45669; A45669.
DR GlycoStatedB; P12839; -.
DR InterPro; IPR006821; Filament_head.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR Pfam; PF04732; filament_head; 1.
DR PRINTS; PR01248; TYPEIKERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
KW Glycoprotein.
FT DOMAIN 0 0 BY SIMILARITY.
FT DOMAIN 1 103 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 844 TAIL.
FT DOMAIN 103 134 COIL_1A.
FT DOMAIN 135 147 LINKER_1.
FT DOMAIN 148 246 COIL_1B.
FT DOMAIN 247 263 LINKER_12.
FT DOMAIN 264 285 COIL_2A.
FT DOMAIN 286 289 LINKER_2.
FT DOMAIN 290 410 COIL_2B.
FT CARBOHYD 47 47 O-LINKED (GLCNAC).
FT /FTId=CAR_000130.

```

FT	CARBOHYD	430	430	O-LINKED (GLCNAC).
FT	MOD_RES	502	502	/FTIG-CAR 000131.
FT	MOD_RES	506	506	PHOSPHORYLATION.
FT	MOD_RES	536	536	PHOSPHORYLATION.
FT	MOD_RES	603	603	PHOSPHORYLATION.
FT	MOD_RES	608	608	PHOSPHORYLATION.
FT	MOD_RES	666	666	PHOSPHORYLATION.
FT	CONFLICT	17	17	MISSING (IN REF. 2).
FT	CONFLICT	21	21	R -> P (IN REF. 2).
FT	CONFLICT	204	204	V -> L (IN REF. 2).
FT	CONFLICT	500	500	MISSING (IN REF. 1).
FT	SEQUENCE	845 AA;	95660 MW;	316C41655B1197D CRC64;

Query Match 18.7%; Score 166.5; DB 1; Length 845;  
 Best Local Similarity 30.9%; Pred. No. 0.11;  
 Matches 60; Conservative 29; Mismatches 80; Indels 25; Gaps 8;

QY	6	EEAEAEAK--RRADAKGKPKGRAGVPGELATPPDKENDAKSSDSVGEETLPSPS	62
DB	559	EEGETEAEAGEGEAEAEAKKEKTEGKVEEMAIKEIKVEPEKAKSPVPKSPVEEVKPKPE	618
QY	63	LKPEKKVAEAEKKVEAEAKKAE-----DQKEDRRNYPTNTYKTLLEIAESDVE--	112
DB	619	AKAGKDEQKEKEKEVEKESPKREKVEKEKPKDVPDK--KKAESPVEKAVEEM	676
QY	113	--VKAELELVKEAKE-PRNEEKVKQAKAVESSKAAEATRLKIKITDRKAE---EE	164
DB	677	ITITSVSVLSLEKDKKEKPOQOEKVE-KAEEGSGSEEVG-DKSPQESKKEDIAINGE	734
QY	165	AKRAAEEDKVKEX	178
DB	735	VEGKEEEEOETOK	748

Search completed: November 21, 2003, 13:32:33  
 Job time : 6.20984 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:37 ; Search time 19.9667 Seconds  
(without alignments)  
1285.946 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537  
Sequence: 1 MFASKSRKRVHYISIRKFSVG.....NRLTQGPPTKPKRPAQSTP 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 261868 seqs, 50246823 residues

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Pending Patents AA New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	1	PCT-US03-27401-468 Sequence 468, App
2	2516	99.2	693	6	US-10-473-928-4598 Sequence 4598, App
3	397.5	15.7	744	1	PCT-US03-27401-251 Sequence 251, App
4	397.5	15.7	744	6	US-10-473-928-32 Sequence 32, App
5	361.5	14.2	1033	6	US-10-258-899A-3455 Sequence 3455, App
6	360	14.2	1026	6	US-10-258-899A-1487 Sequence 1487, App
7	283	11.2	916	6	US-60-490-890-475 Sequence 475, App
8	270	10.6	487	1	PCT-US03-11802-21 Sequence 21, App
9	265.5	10.5	793	7	US-60-490-890-2605 Sequence 2605, App
10	262	10.3	612	7	US-60-500-337-2473 Sequence 2473, App
11	262	10.3	612	7	US-60-500-337-2475 Sequence 2475, App
12	259.5	10.2	8625	6	US-10-679-063-21234 Sequence 21234, App
13	257.5	10.1	4684	1	PCT-US03-14382-432 Sequence 432, App
14	257.5	10.1	4684	1	PCT-US03-14382-434 Sequence 434, App
15	257.5	10.1	4684	1	PCT-US03-14382A-432 Sequence 432, App
16	257.5	10.1	4684	1	PCT-US03-14382A-434 Sequence 434, App
17	257.5	10.1	4684	7	US-60-479-073-497 Sequence 497, App
18	257.5	10.1	4684	7	US-60-479-073-506 Sequence 506, App
19	253.5	10.0	2524	7	US-60-502-656-219 Sequence 219, App
20	253.5	10.0	2524	7	US-60-512-690-253 Sequence 253, App
21	253.5	10.0	4576	1	PCT-US03-31476-281 Sequence 281, App
22	252.5	10.0	4576	1	US-60-479-073-495 Sequence 495, App
23	252	9.9	665	5	US-09-820-843B-107 Sequence 107, App
24	251	9.9	747	6	US-10-425-114A-67803 Sequence 67803, App
25	250.5	9.9	552	7	US-60-500-337-2474 Sequence 2474, App
26	250	9.9	1270	6	US-10-473-040-696 Sequence 696, App

27	250	9.9	2067	6	US-10-473-040-778 Sequence 778, App
28	247	9.7	662	7	US-60-500-337-2476 Sequence 2476, App
29	247	9.7	718	6	US-10-258-899A-3786 Sequence 3786, App
30	245.5	9.7	699	6	US-10-258-899A-1818 Sequence 1818, App
31	244.5	9.6	1435	6	US-10-473-551-661 Sequence 661, App
32	244.5	9.6	1021	6	US-10-473-551-1057 Sequence 1057, App
33	244	9.6	507	6	US-10-425-114A-47285 Sequence 47285, App
34	242.5	9.6	774	6	US-10-258-899A-3849 Sequence 3849, App
35	240	9.5	1533	6	US-10-679-063-23311 Sequence 23311, App
36	234.5	9.2	680	6	US-10-425-114A-68086 Sequence 68086, App
37	232	9.1	8943	6	US-10-679-063-20888 Sequence 20888, App
38	231	9.1	748	6	US-10-425-114A-72422 Sequence 72422, App
39	229.5	9.0	1848	6	US-10-687-046-6 Sequence 6, App
40	226.5	8.9	743	1	PCT-US02-18638A-188 Sequence 188, App
41	225	8.9	1404	6	US-10-473-576-2 Sequence 2, App
42	224.5	8.8	843	6	US-10-258-899A-1144 Sequence 1144, App
43	224.5	8.8	1879	6	US-10-296-115-1265 Sequence 1265, App
44	224.5	8.8	1960	6	US-10-258-899A-1516 Sequence 1516, App
45	224.5	8.8	1963	6	US-10-258-899A-3484 Sequence 3484, App

#### ALIGNMENTS

```
RESULT 1
PCT-US03-27401-468
Sequence 468, Application PC/TUS0327401
GENERAL INFORMATION:
APPLICANT: TUFTS UNIVERSITY
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
FILE REFERENCE: 700355-52941-PCT
CURRENT APPLICATION NUMBER: PCT/US03/27401
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US 60/407,082
PRIOR FILING DATE: 2002-08-10
NUMBER OF SEQ ID NOS: 560
SOFTWARE: PatentIn version 3.2
SEQ ID NO 468
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-468

Query Match 99.2%; Score 2516; DB 1; Length 693;
Best Local Similarity 99.6%; Pred. No. 1.8e-75;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MFASKSRKRVHYISIRKFSVG-ASYVVASLVMGVSVAHTEGATQVPTSSNRANESQAO 59
DB 1 MFASKSRKRVHYISIRKFSVGASVVASLVMGVSVAHTEGATQVPTSSNRANESQAO 60
QY 60 GEPKRLDSDRDARKEVEYVVKIYESYAKSTKSHHTTVLVLNINIKNEYLNKIV 119
DB 61 GEPKRLDSDRDARKEVEYVVKIYESYAKSTKSHHTTVLVLNINIKNEYLNKIV 120
QY 120 ESTSESOLQILMESRSKVDVAVSKEPKDSSSSSSDSTKPEASDTAKNPKTEPEKV 179
DB 121 ESTSESOLQILMESRSKVDVAVSKEPKDSSSSSSDSTKPEASDTAKNPKTEPEKV 180
QY 180 AEAKKVVEEA-KKAKOKEDRRNYPITTKTLELAESDVEVKAELELVKANEPR 238
DB 181 AEAKKVVEEA-KKAKOKEDRRNYPITTKTLELAESDVEVKAELELVKANEPR 240
QY 239 DEQIKQAEAEVSKQAEATRLTKITDREAESEEAERADAEQKPKGRARGVGEL 298
DB 241 DEQIKQAEAEVSKQAEATRLTKITDREAESEEAERADAEQKPKGRARGVGEL 300
QY 299 ATPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAEKVEAEKKAEDQKEEDRRNYP 358
DB 301 ATPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAEKVEAEKKAEDQKEEDRRNYP 360
QY 359 NTYKTELELAESDVEVKAELELVKAEKPEKNEKVKQAKAEVSKKAEATRLKIKT 418
```

```
|||||
Db 361 NNYKLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 420
Qy 419 DRKKAEEBAKRAAEDVKYKPAOPAPAPAKPAKPAKPPENPAEQPAEKPADQ 478
Db 421 DRKKAEEBAKRAAEDVKYKPAOPAPAPAKPAKPAKPPENPAEQPAEKPADQ 480
Qy 479 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 511
Db 481 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 513

RESULT 2
US-10-472-928-4598
; Sequence 4598: Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqMan99, version 1.03
; SEQ ID NO: 4598
; LENGTH: 693
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: choline binding protein A (cbpa)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Feature of note: WY motif
; OTHER INFORMATION: Similar to strain R6 sequence 15904036 (0..E+01)
US-10-472-928-4598
```

Query Match 99.2%; Score 2516; DB 6; Length 693;

Best Local Similarity 99.6%; Pred. No. 1.8e-75; Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```
Qy 1 MFASSEKRVHSIRKFSVG-ASVVVASLWGSVVAATENEGATOPTSSNRANESQAEQ 59
Db 1 MFASSEKRVHSIRKFSVGASVVVASLWGSVVAATENEGATOPTSSNRANESQAEQ 60
Qy 60 GEOPKKLDSEDRKAREVEEYKTVGSYAKSTKRRTITVALYNELNNINNEYLNTKY 119
Db 61 GEOPKKLDSEDRKAREVEEYKTVGSYAKSTKRRTITVALYNELNNINNEYLNTKY 120
Qy 120 ESTSSQQLIMMESRSKYDEAVSKPEKSSSSSSSDSTKPEASDTAKPKNPTPEGEKV 179
Db 121 ESTSSQQLIMMESRSKYDEAVSKPEKSSSSSSSDSTKPEASDTAKPKNPTPEGEKV 180
Qy 180 AEAKKKVEBA-KKAKDQKEDRRNPTTYKTLLELEIASDVVKAAELIVKANER 238
Db 181 AEAKKKVEBAEKKADQKEDRRNPTTYKTLLELEIASDVVKAAELIVKANER 240
Qy 239 DEOKIKQAEAVESKQAEATRLKIKTDRKEAEKEERKRAKAEQKPKRAKRGVPGEL 298
Db 241 DEOKIKQAEAVESKQAEATRLKIKTDRKEAEKEERKRAKAEQKPKRAKRGVPGEL 300
Qy 299 ATPDKKENDAKSSDSVGEETLPSPSLKEPEKVAEAEKKEVEAKKAEQKEDRRNPT 358
Db 301 ATPDKKENDAKSSDSVGEETLPSPSLKEPEKVAEAEKKEVEAKKAEQKEDRRNPT 360
Qy 359 NTYKTLLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 418
Db 361 NTYKTLLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 420
Qy 419 DRKKAEEBAKRAAEDVKYKPAOPAPAPAKPAKPAKPPENPAEQPAEKPADQ 478
Db 421 DRKKAEEBAKRAAEDVKYKPAOPAPAPAKPAKPAKPPENPAEQPAEKPADQ 480
```

```
Qy 479 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 511
Db 481 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 513
```

RESULT 3  
PCT-US03-27401-251

```
; Sequence 251: Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 251
; LENGTH: 744
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-251
```

Query Match 15.7%; Score 397.5; DB 1; Length 744;

Best Local Similarity 28.0%; Pred. No. 6.8e-07; Matches 147; Conservative 89; Mismatches 166; Indels 123; Gaps 22;

```
Qy 47 TSNRANESQAEQEOPEPKLDSEDRKAREVEEYKTVGSYAKSTKRRTITVALVN- 105
Db 63 TAKKAEADQKKYEDDQKTE--EKARKE-----AASQKNDVALVQNA 106
Qy 106 -----ELNNINNEYLNKIVESTSSQQLIMMESRSKYDEAVSKPEKSSSSSDSTK 160
Db 107 YKEYREYQNRQSKY-----KSDAEYQKLT-----TEVDSKIEKARKQODLQNKFEVR 155
Qy 161 PEASDTAKPNPTPEGEKVAAKKKVEAKKAKQKEDRRNPTTYKTLLELEIASDV 220
Db 156 --AVVPEPN-----ALAEYKKAABAKA--EEVAKKRY--DYATILKVALAKKEV 200
Qy 221 EVKKAELIVKANERPEDEOKIKQAEAVESKQAEATRLKIKT-----KT 265
Db 201 EAKLEIEIKL-----QYEISTLEGEVATAGHQVNDLKLKLAGADPDQTEVIEAKL 251
Qy 266 DREAEBAKRAKAEQK-----PKRAKRGV--PGLATPPKENDAKSSSS 314
Db 252 KKGAEELNAKQAEIAKQTELEKLDLSLDPGKTQDELDKAEBAELDKADELQNKVAD 311
Qy 315 VGEETLPSPSL-----KPEKVAE-----AEKKEVEAKKAEQKEDRRNPTTYKTL 365
Db 312 LEKEISNLEILLGADPEEDDTALQNKLAAKKAEIAKQTELEKLDLSLDPGKTQDELD 371
Qy 366 LEIASDVVKAAELE-LVKEAKEPRN-----EEKVKQAQAEVSKKAEATLE 414
Db 372 KEAEBAELDKADELQNKVADLEKEISNLEILLGADSEDDTALQNKLAKEA--LE 428
Qy 415 KITDRKKAEEBAKRAAEDVKYKPAEQ-QAPAPAKPAKPAKPPENPAEQPAEK 473
Db 429 KQKELDLNALNELSPDGEDE--TPAPAPQEQCAPAPAPKPAKPPENPAEQPAEK 486
Qy 474 PADQAEEDYARSEEEYNRLTQOQPPKTEKPAQ-----PSTP 511
Db 487 PAPAPKPPQAPAPKPE-----QPAKPEKPAEPTQPEKPAEP 524
```

RESULT 4  
US-10-472-928-32

```
; Sequence 32: Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
```

Query Match	15.7%	Score 397.5;	DB 6;	Length 744;
Best Local Similarity	28.0%	Pred. No. 6.8e-07;		
Matches 147;	Conservative 89;	Mismatches 166;	Indels 123;	Gaps 22

QY	47	TSNRANESQEQGQPKLUSEBRKARKEVEYVKI VGE5YASTKSRHITVALVN	105
Db	63	TAKKAEADQCKVEYEDDKRTE--EKARE-----AASQLDNALVALVONA	106
QY	106	-----ELNNIKNEYLNKIVESTSESQOLIMMESKRYKDEAVSKPEKSDSSSSSDSSTK	160
Db	107	YKEYREVONQNSKY-----KSDAEQCKL-----TEVDSKIEKARKQODLQNFNEVR	155
QY	161	PEASDTAKPNKPTPEBGEKVAEAKKKVEBAKAKQODEDRNYPITYKTELEIAESDV	220
Db	156	--AVVPEPN-----ALAEKTKKAEBAKA--SEKVAKRKY--DYATLKVALLAKREV	200
QY	221	EKKVAELELVKVNKANEPRDEOKIKQAEAEVSKQAEATRIKKI-----KT	265
Db	201	EAKELIEKL-----QYEISTLEOVATIAQOVNKLKLILAGAPDDGTEVIEAKL	251
QY	266	DREBAEEBAKRRADAKEQGK-----PKGRAKRGV--PGLATPPKENDAKSSDS	314
Db	252	KKGAEALNAKQAEKLKQTELEKLLDSLDPSGKQODELDKEAEALDKKADDELQNKYAD	311
QY	315	VGEETLSPSCS---KPEKKVAE---AEKKVEBAKKAEDQKEEDRNPNTNYKTLE	365
Db	312	LEKISINLEILILGADPEDDTAALQNKLAAKKAELAKQTELEKLLDSLDEPGKQODELD	371
QY	366	LEINSEVEYKKALE-LYKEBAKEPN-----EKKVQAKAVESKKAATLE	414
Db	372	KEAEBAELDKADELQNKVADLEKISINLEILILGADSEDDTALQNLATKKAE--LE	428
QY	415	KIKTRDKKAEBAKKAEEADRVKKEPAEQP-QAPAPAKAEKPAAPKPNPAEQPKAEK	473
Db	429	KTQKELDALNELGPDGDEE--TPAPAPQEQGAPAKKPPQAPAPKPPQAPAPKPEQ	486
QY	474	PADQQAEBDYARRSEEEYNRLTQOQPPKTEKPAQ-----PSTP	511
Db	487	PAPAPKPEQAPAPAPKPE-----QPAKPEKPAEBPTQPEKPAFP	524
RESULT 5			
US-10-258-899A-3455			
Sequence 3455, Application US/10258899A			
GENERAL INFORMATION:			
APPLICANT: Tang, Y. Tom			
APPLICANT: Liu, Chenghua			
APPLICANT: Drmanac, Radoje T.			
APPLICANT: Asundi, Vinod			
APPLICANT: Zhou, Ping			
APPLICANT: Xu, Chongjun			
APPLICANT: Cao, Yicheng			
APPLICANT: Ma, Yungqing			
APPLICANT: Zhao, Qing A.			
APPLICANT: Wang, Duntui			

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	14.2%;	Score 361.5;	DB 6;	Length 1033;
	Burst Local Similarity	26.1%;	Pred. No. 1.2e-05;		
	Matches 155;	Conservative 82;	Mismatches 234;	Indels 123;	Gaps 24;
Oy	5 KSERKVHYSIKTSVGSVVV-----ASLVGVSVAHTENEGATQVPTSSNRANESQAQ 59				
Dd	44S KSEKI--KVVEKSEKTIVIEBQTERTQYTEETVEDKAEKBECKEGBEEBAEG 502				
Oy	60 GEOPKLDSDERDA--RKEVEEYVK-KIYESSTAKSYKKKHHTTVVALYNELNIIKYEYL 116				
Dd	503 GEEETKSPPAEEAASPPEKAESPVEKEAKSPAAKASEKEBAKSPAVKSPPEKAKSP-- 559				
Oy	117 KIVESTESQLILMMRSKVDNAVSKFEKDSSSSSSDSTPKPEADPAKNKP---- 172				
Dd	560 --AKEAKSP-----PAAKSPKDGONFPAEVYKSPKAKSPAKKEAKSPAEEKSPKAK 612				
Oy	173 ---TEPEKVAEAARKV--EEAKKAKDQKEEDRRNPITYTYLTLELTAESDVEYKAELE 228				
Dd	613 SPVKEAKSPAEEKASPVEKEAKSPAKEYSPPEKAKSPT--KEAKSPPEKAKSPPEKAKSP-- 669				
Oy	229 LVKYKANEPDEQIKOALEAVEFSK-----QAAATRLAKTKTD-RBEABEAKRADAKE 282				
Dd	670 --KEEAKSPKAKSPVAAEKASKPEKAKSPVKAALAKSPKAKSPVKEAKSPPEKAKSPVKE 727				
Oy	283 QGRPKGRAGRVGCELATPDKKENDAKSDSSVGEEITLPSLSLKPEKVAEAKEAYEAK 342				
Dd	728 EAASPEKAKSPVVEEAKTPEKAKSPVKEAKS-----PEKAKSP--KAKTUDVKSPEAK 780				
Oy	343 KAAEDQKEEDRRNPITYTYLTLEL-----AESDV--EYKAELEL-VKEAKEPRNEE 394				
Dd	781 TPKEEARSPADFPPEKAKSPVKEEVYKSPKAKSPAKEKAAPKEKIPKKEBYVSGPVKE 840				

Qy 395 -----KVQAKAEVESKKAENTRLKIKTRKKAEEAKRAAEEDVKYK----- 440  
Db 841 EKQEVKKYKPPKAEKBEKAPATPTKEKDSK--EEAPKKEAPKPVKEEKBEAPVEXP 898  
Qy 441 -----PAEQOPAPAPKAEK-----APAPKPPNPA 466  
Db 899 KESKVEAKKEAEEDKKVPTEPEKAPAVEVEDAKPKYKTEVAKKEPDDAKAKEPKPA 958  
Qy 467 EOPKAEKADQQAEDVARRSEEEYNRLTQOQPKTEKPA-----OPSTP 511  
Db 959 EKKEA-APEKDOTKEKAKKPEEK-----PKTEAKAKEDDKTLSKEPSKP 1002

## RESULT 6

US-10-258-899A-1487  
Sequence 1487, Application US/10258899A  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Drmanac, Radoje T.  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhou, Ping  
APPLICANT: Xu, Chongjun  
APPLICANT: Cao, Yicheng  
APPLICANT: Ma, Yunqun  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Wang, Zhi Wei  
APPLICANT: Xue, Aidong  
APPLICANT: Yang, Yonghong  
APPLICANT: Weinman, Tom  
APPLICANT: Goodrich, Ryle  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 787CIP2-26/US  
CURRENT APPLICATION NUMBER: US/10/258,899A  
CURRENT FILING DATE: 2003-11-06  
PRIOR APPLICATION NUMBER: PCT/US01/04098  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: 09/774,434  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 09/728,422  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: 09/693,325  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 09/663,561  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 09/654,936  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 09/620,325  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/598,075  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/560,875  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 09/496,914  
PRIOR FILING DATE: 2000-02-03  
NUMBER OF SEQ ID NOS: 3960  
SOFTWARE: Custom  
SEQ ID NO 1487  
LENGTH: 1026  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-258-899A-1487

Query Match 14.2%; Score 360; DB 6; Length 1026;

Best Local Similarity 26.1%; Pred. No. 1.4e-05;  
Matches 154; Conservative 81; Mismatches 238; Indels 118; Gaps 23;

Qy 5 KSEKRVHSIRKSVGASVV-----ASLVNGSVVHATENEGATQVPTSSNRANSQAEQ 59

Db 439 KSEKRI--KVKEKSEKTEIVIEQTEETQVTEEVTEEEKAEKEEGEGEEBAEG 496  
Qy 60 GEQPK-----LDSEDKARKEVEEVKKIVGSSVAKSTKKRHTTIVALVNLNININE 113  
Db 497 GEEETKSPVVEEASPEKEKSPVKEAK--SPAEEKSPKEBAKPAEVKSEKAKSP 553  
Qy 114 YLNKIVESTESQIIMMSRSKVDAVSGFEKDSSSSSSDSTPEASDTAKPNKPT 173  
Db 554 -----AKEBAKSPPEAKSPKEKAPAEVYKSPPEKAKSPAKEKAPAEKSPPEKAKSPV 608  
Qy 174 -EPGEKVAEAKKV-EEAKKADQKEEDRRNPTTYKTLELTAESDVEYKAELELVK 231  
Db 609 KEEAKSPAEKAPKPEKAPAEVYKSPPEKAKSP-KEEAKSPPEKAKSPPEKAKSPPEKAKSP 663  
Qy 232 VKANEPPEQKIKQAEVEVSK-----QAEATRLKIKTD-REAEAEARRADAKQCK 285  
Db 664 EEAKEKSPPEKAKSPVKEKSPPEKAKSPVKEKAKSPPEKAKSPPEKAKSPPEKAKSP 723  
Qy 286 PKGAKRGVPGELATPPKENDAKSDSVGEETLPSPSLKPEKVAEAEKVEAKKA 345  
Db 724 SPEKAKSPVKEAKTPPEKAKSPVKEAKS-----PEKAKSP-KATLDVKSPEAKTPA 776  
Qy 346 EDQKEEDRRNPTTYKTLEL-----AESDV--EVYKAELEL-VKEAKEPNEB--- 394  
Db 777 KEEARSPADKPEKAKSPVKEEVKSPPEKAKSPLEKADAKPEKEIPKKEEVKSPVKEBEKP 836  
Qy 395 ---KVQAKAEVESKKAENTRLKIKTRKKAEEAKRAAEEDVKYK----- 440  
Db 837 QEVYKPPKAEKBEKAPATPTKEKDSK--EEAPKKEAPKPVKEEKBEAPVEXP 894  
Qy 441 -----PAEQOPAPAPKAEK-----APAPKPPNPAEQ 469  
Db 895 KVEAKKEAEEDKKVPTEPEKAPAVEVEDAKPKYKTEVAKKEPDDAKAKEPKPAEK 954  
Qy 470 KAEKPADQQAEDVARRSEEEYNRLTQOQPKTEKPA-----OPSTP 511  
Db 955 EA-APEKDOTKEKAKKPEEK-----PKTEAKAKEDDKTLSKEPSKP 995

## RESULT 7

US-60-490-890-475  
Sequence 475, Application US/60490890  
GENERAL INFORMATION:  
APPLICANT: Li, Martha  
APPLICANT: Rudnow, Brent A.  
APPLICANT: Webster, Kevin R.  
APPLICANT: Jackson, Donald  
APPLICANT: Wong, Tai W.  
TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION  
FILE REFERENCE: D0310 PSP  
CURRENT APPLICATION NUMBER: US/60/490,890  
CURRENT FILING DATE: 2003-07-29  
NUMBER OF SEQ ID NOS: 2779  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 475  
LENGTH: 916  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-490-890-475

Query Match 11.2%; Score 283; DB 7; Length 916;

Best Local Similarity 24.0%; Pred. No. 0.0039;  
Matches 118; Conservative 85; Mismatches 199; Indels 90; Gaps 16;

Qy 53 NESQAEQEQPKLDSERDARK-----EVEEYKKIVGESYAKST-----KKR 96  
Db 368 NELRGTKEMARHLREYQDLLNVYGMALDIAAYRKLLBGEFFRSTFAGSITPPLYTHR 427  
Qy 97 HTTIVALVNLNIN-----NIKIEYLNKIVESTESQIIMMSRSKVDAVSGFEKDS 149  
Db 428 PPIITSSKIOKTKVEAPKLVQHKFVEIIEETKVEDEKSEMEBALTAITEELAAKKEE 487

Qy 150 SSSSSDSTYKPEASD---TAK--ENKTEPGEKVAEAKKVEEAKKADQKEEDRRNP 204  
Db 488 KKEAAEEKEEPEEAEVEEAAKSPVKATAPVEKEEGBEGEEDBEGACAS-- 545  
Qy 205 TTTTLELELAESDVEVYKAALELVKYNAPRDEQIKOAEAVESKQAEATRLKIK 264  
Db 546 -----DOAEGBGSEKESSE-----KEEGBEGEGETEAEABEAEKKEKVE 589  
Qy 265 TDREAEAEARPADAKGQKRGKRGKRGVDELATPDKKENDAKSSSSVGEETLPSPS 324  
Db 590 EKSEEVATKEELVADAKVE--KPE--KAKSPVPKSPVEEKGKSPVPKSPVEEKGKSPVPSP 647  
Qy 325 LKPEKVAEAEKVEE--AKKTAEDQKEEDRRNPNTYKTLLELAESDVEVYKAALELV 383  
Db 648 VEEKGKSPVPKSPVEEKGKSPVPKSPVEEKGKSP-----VPKSPVEEAKSKAEVG 697  
Qy 384 KEAEKPEEKEKVAEAKKVEESKAEATRLKIK--TURKAEAEAKKAA----- 432  
Db 698 KGEQKE--EEKEKVEKAPKEEKEKKE-----EKPKDVPKKAESPVEKEAVAEVYITKS 752  
Qy 433 -----EEDKYKEKPAEPOPAPAPKAPKAPKPEPAPQPAKAKADQAEEDVARR 486  
Db 753 VKVHLEKTEKEEGPLQ-----EKEKEKAGGEGSEEGSDKAGKSKREDAVAVN 803  
Qy 487 SEEEYNRLTQOQ 498  
Db 804 GEVEGKEVEGE 815

RESULT 8  
PCT-US03-11802-21  
Sequence 21, Application PC/TUS0311802  
GENERAL INFORMATION:  
APPLICANT: Washington University  
TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protective  
FILE OF INVENTION: Immunogenicity  
FILE REFERENCE: 56029-40434  
CURRENT APPLICATION NUMBER: PCT/US03/11802  
CURRENT FILING DATE: 2003-04-15  
PRIOR APPLICATION NUMBER: US 60/373,626  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/372,616  
PRIOR FILING DATE: 2002-04-15  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 21  
LENGTH: 487  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
PCT-US03-11802-21

Query Match 10.6%; Score 270; DB 1; Length 487;  
Best Local Similarity 28.6%; Pred. No. 0.0069;  
Matches 127; Conservative 66; Mismatches 165; Indels 86; Gaps 23;  
Qy 119 VESTSESOLOLIMMESKVDVAIS--KFEKSSSSSSDSTKPEASDTAKPNKPTERGE 177  
Db 1 LQANESQK---EADKTKKATQKDEAAAPATITTTTIVPEPSELAKTKKAEAT 56  
Qy 178 KVAE-AKKVVEA-----KKAQKEEDRRNPITTYKTL-----EIASD 219  
Db 57 KEAAVAKKSEEAKEVEKKNKTLLEDAENKIKDIVLQNKVALLKGIAPYQNEVAFELN 116  
Qy 220 VEVKAALELVKYNAPRDEQIKOAEAVESKQAEATRLKIKITDREAEAE 273  
Db 117 KEIARLOSLEDAENNVEDYIEGLEQAITNKKAELATQ---QNDIKQKLEDAELE 173  
Qy 274 AKR--RAAKGEGKRGKRA--KRGVPEL-----ATPDKKENDAKSSSVGEETLPSPS 322  
Db 174 LEKVLATLDPGKQODELDEKAAEALNEKVEALQNOVALEELSLKLENDLKDAETLOS 233  
Qy 323 P---SLAPEKVAEAEKVEEAKKAEAD--QKEED-----RNPYPTNYKTL---LEIAE 370

Db 234 PVASQSAEKDYDAKADKAKKAVEDAQALDDAKAAQKKYEDDQKTEEKALEKRA 293  
Qy 371 SDVEVKAALELVYEE--AKEPNEEKVQKAAEV--ESKAEATRLKIKITDR----- 420  
Db 294 SE-EMDYA--VAVQAVIATQOATDKAKADADKXIDAEKAREEAKTKFNTVAMVPE 351  
Qy 421 -----KKAEEAEKRA-----AEEDKYKEKPAEPOPAPAPK--AEKPAAPK--P 462  
Db 352 PEQIAETKKSSEAKQKAPETLTKLEBAKALDEAEKATKATAKQVDAEVAPOAKIAEL 411  
Qy 463 ENPAEQKAE--KPADQOAEEDYAR 485  
Db 412 ENQVHRLQELKEIDESESEEDYAK 435

RESULT 9  
US-60-490-2605  
Sequence 2605, Application US/60490890  
GENERAL INFORMATION:  
APPLICANT: Li, Martha  
APPLICANT: Rupnow, Brent A.  
APPLICANT: Webster, Kevin R.  
APPLICANT: Jackson, Donald  
TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION  
FILE REFERENCE: D0310 PSP  
CURRENT APPLICATION NUMBER: US/60/490,890  
CURRENT FILING DATE: 2003-07-29  
NUMBER OF SEQ ID NOS: 2779  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 2605  
LENGTH: 793  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-490-890-2605

Query Match 10.5%; Score 265.5; DB 7; Length 793;  
Best Local Similarity 22.2%; Pred. No. 0.013;  
Matches 123; Conservative 105; Mismatches 192; Indels 135; Gaps 23;  
Qy 53 NESQAEQEQPKLDSERDKARKEVEEYKVIKIGESYAKSTKRRHTTTVALVNLNIN 112  
Db 30 NDDDEEBAERERRRARRQERLQKQEE-----ESLGQVTDQ-----VEVNAQNSVPD 76  
Qy 113 E-----YLNKIVSTSSQI--QILMESR--SKVDVAVKFEK-----DSSS----- 152  
Db 77 EAKTTTNTQVGGDDAEALERLARREERRQKRLQELERQKEFDPITTDASISLPSRR 136  
Qy 153 -----SSSDSTKPEASDTAKPNKPTERGEKVAEAKK-----VEEAKKAKDQKEEDRRN 202  
Db 137 MQNDTAENETTEKESESQERYEIEETLVTKSYQKNDWRDAEENKEDKEKEE 196  
Qy 203 YP---TTTYTLELEI-----AESDVEYKAALELVKYNAPRDEQIKOAEAEV----- 250  
Db 197 KPKRGSIGENQVVMVEKTESEQETVVMSLKNGQISSEPEKQOEEREGSDLSHHEK 256  
Qy 251 -----ESKQAEATRLKIKITDREAEAEAK-----RADAKGEGKRGKRGVPEELA 299  
Db 257 MEEDKEKRAAEARLAEERERIKAEQDKIADERARIAEKKAAQOERERREAE--ERE 315  
Qy 300 TPDKENDAKSSSVGEETLPSPSLKPEKVAEAEKVEEAKKAAEDQ-----KEEDRN 355  
Db 316 RMREBEKRAAEERORIEE-----EKRAAEERRIEERERRAAEERORIEEERKRA 366  
Qy 356 YPTNYKTLLELAESDVEVYKAALELVKEAEKPEEVEKVAQKAAE-----VESKKA 408  
Db 367 AEERQARAE--EEKAKVEEGKRNKQI--EKKRAVQETIKIKEEKVQKIEGKVVNKKKA 423  
Qy 409 EATRL-----KITDRKAAEAEKAKKAAEDK-----VK--EKPAEPOPAPAPKA--- 453  
Db 424 QEDRLQTAVALKQGEKGTVAQAKREKLQEDKPTFKKEIKDKKIKDKPEKVSFMD 483  
Qy 454 -----EKPAAPKPEENPAEQKAEKPAQOAEEDVARR 486

Db 484 RKKGFTEVKSQNGEFMTHTKHTENTFSRPGSRASVDPTKEAGAPQVBAKRLBELRRR 543  
OY 487 SE---EFTNRLTQOO 498  
Db 544 GETESEFEKLOKQ 558

## RESULT 10

US-60-500-337-2473  
; Sequence 2473, Application US/60500337  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01483  
; CURRENT APPLICATION NUMBER: US/60/500,337  
; CURRENT FILING DATE: 2003-09-05  
; NUMBER OF SEQ ID NOS: 123188  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2473  
; LENGTH: 612  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-500-337-2473

Query Match 10.3%; Score 262; DB 7; Length 612;  
Best Local Similarity 23.9%; Pred. No. 0.014;  
Matches 124; Conservative 80; Mismatches 197; Indels 118; Gaps 19;

OY 55 SOAEGEQPKLDSERDKARKEVEEYVKIYESYAKSTKKRHTITVALVNLNNIKRY 114  
Db 5 SOKAEIKEMLASDDEEDVSSKVEKAYVPLTG---TVKGRFA-----EMEKQROE 52  
OY 115 LNKIVESTESQLOILMESRSKVDVAVSKEKSSSSSSDSTKPEASD---TAKPN 170  
Db 53 QKRTTEERKRRIODMLEKRIQRELAQRAQIEDINTGTESASEGDDSLITVVPV 112  
OY 171 KPTPEGEVAEAKKVEAKKADOK-----EEDRRNYPTITY-----KTLLEI 215  
Db 113 KSYTSGM---KNFEDLEKEREKERIKYEEDR---IRYEOPLSLKCAKCLSLVM 165  
OY 216 AEDSVEYKKALELVYK-----ANPRDOKIKQAEV-----ESKQAEATRLKI 263  
Db 166 DDEIESEKKSLSLPGKLTFFELERQRENKQAEERKRLIEEKKAFAEARRQMV 225  
OY 264 KTDREAE-----EAKRRADAKQCKPKGRKGVPGE-LATPDKE 305  
Db 226 NEDENDOTAKIFKGYRPGKLTLSFEEMERQREDEKKAEEARRRIIEEKKAFAEARR 285  
OY 306 NDAKSSDSSVGEETLPSPSLKPEKVAEAE---KVEEAKKADQKEDRRNYPTINTY 361  
Db 286 NMVVDSDSPENYKTIQSEFLTTPGKLEINFELLKQMEERKRTEERKH----- 335  
OY 362 KTLLEIAESDVEYKKALELVKEAKEPRN-----EKKVQAKAEVSKKAA 410  
Db 336 -----KLEMEKQFEQOLRQEMGEEBEENETFGLSREYBELIKTKRSGSIQAKLK 385  
OY 411 TRLEKI-----KTRKKAEEBAKRAAEEDKVEKPAE-----QPPAPAPKAEKAPA 459  
Db 386 SKFEKIGLSEKEIOKRIEERARRAIDLEIKEREANFHEEDVDVPRARKSE---A 441  
OY 460 PKPENPAEQPAKPAEQADQAEEDYARSEEEYNRLTQOO 498  
Db 442 PFTHKVMKARFEQMAKAREEEOGRIBEOQLLRMOFEQ 480

RESULT 11  
US-60-500-337-2475  
; Sequence 2475, Application US/60500337  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01483  
; CURRENT APPLICATION NUMBER: US/60/500,337  
; CURRENT FILING DATE: 2003-09-05  
; NUMBER OF SEQ ID NOS: 123188  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2475  
; LENGTH: 612  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-500-337-2475

Query Match 10.3%; Score 262; DB 7; Length 612;  
Best Local Similarity 23.9%; Pred. No. 0.014;  
Matches 124; Conservative 80; Mismatches 197; Indels 118; Gaps 19;

OY 55 SOAEGEQPKLDSERDKARKEVEEYVKIYESYAKSTKKRHTITVALVNLNNIKRY 114  
Db 5 SOKAEIKEMLASDDEEDVSSKVEKAYVPLTG---TVKGRFA-----EMEKQROE 52  
OY 115 LNKIVESTESQLOILMESRSKVDVAVSKEKSSSSSSDSTKPEASD---TAKPN 170  
Db 53 QKRTTEERKRRIODMLEKRIQRELAQRAQIEDINTGTESASEGDDSLITVVPV 112  
OY 171 KPTPEGEVAEAKKVEAKKADOK-----EEDRRNYPTITY-----KTLLEI 215  
Db 113 KSYTSGM---KNFEDLEKEREKERIKYEEDR---IRYEOPLSLKCAKCLSLVM 165  
OY 216 AEDSVEYKKALELVYK-----ANPRDOKIKQAEV-----ESKQAEATRLKI 263  
Db 166 DDEIESEKKSLSLPGKLTFFELERQRENKQAEERKRLIEEKKAFAEARRQMV 225  
OY 264 KTDREAE-----EAKRRADAKQCKPKGRKGVPGE-LATPDKE 305  
Db 226 NEDENDOTAKIFKGYRPGKLTLSFEEMERQREDEKKAEEARRRIIEEKKAFAEARR 285  
OY 306 NDAKSSDSSVGEETLPSPSLKPEKVAEAE---KVEEAKKADQKEDRRNYPTINTY 361  
Db 286 NMVVDSDSPENYKTIQSEFLTTPGKLEINFELLKQMEERKRTEERKH----- 335  
OY 362 KTLLEIAESDVEYKKALELVKEAKEPRN-----EKKVQAKAEVSKKAA 410  
Db 336 -----KLEMEKQFEQOLRQEMGEEBEENETFGLSREYBELIKTKRSGSIQAKLK 385  
OY 411 TRLEKI-----KTRKKAEEBAKRAAEEDKVEKPAE-----QPPAPAPKAEKAPA 459  
Db 386 SKFEKIGLSEKEIOKRIEERARRAIDLEIKEREANFHEEDVDVPRARKSE---A 441  
OY 460 PKPENPAEQPAKPAEQADQAEEDYARSEEEYNRLTQOO 498  
Db 442 PFTHKVMKARFEQMAKAREEEOGRIBEOQLLRMOFEQ 480

RESULT 12  
US-10-679-063-21234  
; Sequence 21234, Application US/10679063  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52054)B  
; CURRENT APPLICATION NUMBER: US/10/679,063  
; CURRENT FILING DATE: 2003-10-02  
; PRIOR APPLICATION NUMBER: 60/415,758  
; NUMBER OF SEQ ID NOS: 27373  
; SEQ ID NO 21234  
; LENGTH: 8625  
; TYPE: PRT  
; ORGANISM: Procamburus clarkii  
US-10-679-063-21234



Query Match 10.2%; Score 259.5; DB 6; Length 8625;  
Best Local Similarity 22.9%; Pred. No. 0.092;  
Matches 119; Conservative 84; Mismatches 200; Indels 117; Gaps 17;

QY 77 VEEVYKTVIGESVAKSTK-----RHITVALVNLNNIKYLNKIVE 120  
DB 659 VSVFNRKVVIECVVISTSPACSWFKDSSKVLDRHCV-----NINEGGKVV- 708  
QY 121 STSQQLOIIMWESSKVD-----EAVSFKEKSSSSSSDSTKEASDTAKPKPTPG 176  
DB 709 ---VQWMLKAE---KWDGSGYKVAALNKGKETTFAARGAHHRMLMEKEKKK----- 756  
QY 177 EKVAEAKKVEEAKKADQKEEDRRNYPTITYKTELEIASDVEYKAELELVKANE 236  
DB 757 EEVEEKEKQEQEAAQVEEESKKEEAVKDKPKREIKKEAPVDFTVOLKTKYKPKKEE 816  
QY 237 PRDE-QIKQAEAEVESKQA---EATRLKKIKTDREAEAEAKRADAKEQKPKGRK 291  
DB 817 AADBAQQIETLKKIIPDKEQQAIVLEGKVKLKKTAKAEELTIEAQVKVQLQKGAKEQQAIVE 876  
QY 292 RGVGELATPDKENDAKSSDSSVGEETLPPSPSLKPEKKVAEAEK-KVEEAKKAAEDQKE 350  
DB 877 EGEVYKTKKVAKEEQATEEAQKV---OLKRVGAKEQQAIVEEGEKVKLKVRAR-EQAT 932  
QY 351 EDRRNYPTNTYKTELEIASDVEYKAELELVKEEAKEPNNEEKVKQ--AKAEVESKKA 408  
DB 933 EEAQVKQKVGAEQQAQVEGEKVYKLVKGAKEEQAVEBEPRAKLVKVGARKEAVEEA 992  
QY 409 EATRLKKIKTDKKAEEBAKRAAEEDKVKERPAEQOPAPAPRAE-----454  
DB 993 EKVOIKVEIYKVAKEEAEAEETKQESLQKPKIYE-PRSDAPKIEVLIKDSPPRRSLA 1051  
QY 455 -----KPAAPKPEPNA-----EQKAE-----472  
DB 1052 TGSQPPRRGSLVPPPEGAGRRPSLLIADBEQKTRPGEBAVEFKGKRLRPGSELVGG 1111  
QY 473 ----KPADQQAEEYARSEEYARLQQQPPKTEKPAQ 508  
DB 1112 KGKTKPGEQATKQ---RRRPSDVARPPSVQDDERKMDKCTP 1149

RESULT 13  
PCT-US03-14382-432  
; Sequence 432, Application PC/TUS0314382  
; GENERAL INFORMATION:  
; APPLICANT: Emory University  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING  
; FILE REFERENCE: 12804-011W01  
; CURRENT APPLICATION NUMBER: PCT/US03/14382  
; CURRENT FILING DATE: 2003-05-07  
; PRIOR FILING DATE: 2002-05-07  
; NUMBER OF SEQ ID NOS: 501  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 432  
; LENGTH: 4684  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US03-14382-432

Query Match 10.1%; Score 257.5; DB 1; Length 4684;  
Best Local Similarity 20.6%; Pred. No. 0.073;  
Matches 117; Conservative 125; Mismatches 196; Indels 129; Gaps 21;

QY 40 EGATQVPTSSNRANES-----QAEQGEQPKKL-DSEEDKARKVEEYKTVIGESVAKST 93  
DB 1763 EAERQLERMQLKANLRLQLAEVYLQKSLAQAEAEKQKEAEERARR-RGAKEQAV 1821  
QY 94 KKGHTTVALVNLNNIKYLNKIVESTSESQI---QILMM-----ESRSKVDEA 141  
DB 1822 RQR-----ELAQELEKQRLAEGTAQORLAAGELIRLRAETEGEQQRQLLEE 1872

QY 142 VSKFEKSSSSS-----SSDSTKEASDTAKPKNP-----172  
DB 1873 LARLQREAAATQYRQELAEELAVRAEMEVLLASKAAEESSTSEKSKQRLAEAGR 1932  
QY 173 -TEDEKVAEAKKVEEAKKADQKEEDRRNYPTITYKTELEIA-----ESDVE 221  
DB 1933 PRELAEEARLRALAEAEKQRLQAEEDDANQRAEAEVLAEXLAALIGENTRLKTEAIA 1992  
QY 222 VKKAELELVKVA-----NEPDEOKIQAEAEVESKQAEATRLKKIKTDREE--AE 272  
DB 1993 LKEKAEAEERLRLAEDEAFQRRRLIEQAAQHKADIIEIRLQILKASPSLEERQKGLVED 2052  
QY 273 EAKRADAKEQ-----GKPK-----GRARGVGEATPDKENDAKSSDSS 314  
DB 2053 TLRRQVVEETLALKASFEEAAGKAELELELIRISNADTLRSKEQAELEAARQOL 2112  
QY 315 VGEETLPPSLKPEKKVAEAEKVEEAKKKAEDQKEEDRRNYPTNTYKTELEIASDVE 374  
DB 2113 AAE-----ERRRREAEERVQ--KSLAEAEARQRAALEVEERLKAKEAEARSL 2161  
QY 375 VKKAE-----LELVKEA--KEPRNEK-----VKQAEVESKKAETRLKIKTDREK 422  
DB 2162 RERAEQESARQLQIAQEAQKRLQAEKKAFAVQOEQEIQ---QTLQEOGVLDRLR 2217  
QY 423 AEEBAKRAAEEDKVKERPAEQP-----QAPAPKAEKPAKPEPNAEQPAKPAD 476  
DB 2218 GEAPARRAEAEAEAEAVQAEAREASRRQVEEAEERLKQSEBOQAAPAOAAAEK-LR 2276  
QY 477 QQAEEYARRSEEEYARLQQQPPKTE 503  
DB 2277 KEAQEAEARRAQAQEAALRQKQAADAE 2303

RESULT 14  
PCT-US03-14382-434  
; Sequence 434, Application PC/TUS0314382  
; GENERAL INFORMATION:  
; APPLICANT: Emory University  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING  
; FILE REFERENCE: 12804-011W01  
; CURRENT APPLICATION NUMBER: PCT/US03/14382  
; CURRENT FILING DATE: 2003-05-07  
; PRIOR FILING DATE: 2002-05-07  
; NUMBER OF SEQ ID NOS: 501  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 434  
; LENGTH: 4684  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US03-14382-434

Query Match 10.1%; Score 257.5; DB 1; Length 4684;  
Best Local Similarity 20.6%; Pred. No. 0.073;  
Matches 117; Conservative 125; Mismatches 196; Indels 129; Gaps 21;

QY 40 EGATQVPTSSNRANES-----QAEQGEQPKKL-DSEEDKARKVEEYKTVIGESVAKST 93  
DB 1763 EAERQLERMQLKANLRLQLAEVYLQKSLAQAEAEKQKEAEERARR-RGAKEQAV 1821  
QY 94 KKGHTTVALVNLNNIKYLNKIVESTSESQI---QILMM-----ESRSKVDEA 141  
DB 1822 RQR-----ELAQELEKQRLAEGTAQORLAAGELIRLRAETEGEQQRQLLEE 1872  
QY 142 VSKFEKSSSSS-----SSDSTKEASDTAKPKNP-----172  
DB 1873 LARLQREAAATQYRQELAEELAVRAEMEVLLASKAAEESSTSEKSKQRLAEAGR 1932  
QY 173 -TEDEKVAEAKKVEEAKKADQKEEDRRNYPTITYKTELEIA-----ESDVE 221  
DB 1933 PRELAEEARLRALAEAEKQRLQAEEDDANQRAEAEVLAEXLAALIGENTRLKTEAIA 1992

```

OY      222 VKKALTELIVKYVA-----NEPPEOKITKOAEEVESKOAPATYKIKITDREE--ABE 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1993 LKEKAENEBRLRLAEDFAFORRLIEEBOAQHKADITIEERLAIOTKRASSSELEKRGVLVED
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      273 EAKRRADAKEO-----GKPK-----GAKRGVPELATPDPKENDAKSSSS 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2053 TLRQRRQYVEEETLALKASPEKAAAKALTELELGRIRSNABEDTLASKGOALHEAARQOPL 2112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      315 VGEETLPSBPLKPEKKVAEAEKKVEEAKKKAEDQKEEDRNNPTNTYKTELEIEASDVE 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2113 AAEE-----ERRREAEERVO--KSLAAEEEAARQKALTEVERLTKAIVEARSL 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      315 VKKAE-----LELYKEEA--KEPPEEEK-----VKOAAEVESKKAELATRLKITDRKK 422
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2162 RERAEOSARQIOTLOEAOAKOKILOAEFEKKAHAFVAOQKEOLO---QTLLOEOSVLDBLR 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      423 ABEKKRKAABEDKYKEKPAEDP-----QAPAKAEKPAKPAKPEKPAEDPKAEKPAD 476
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2218 GEAEKARRPAAEEAEARVQAEEREAQAOSRRQYVEAEERLTKOSAEEOQAORQAQAAMAEK--LR 2276
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      477 OOAEDVARRSEEEYNRLTQOOPPKTE 503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2277 KEAEQEARRRAQAEQALRQKQAADAE 2303

```

```

RESULT 15
PCT-US03-14382A-432
; Sequence 432, Application PC/TUS0314382A
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: ANTI-RETROVIRAL AGENTS
; FILE REFERENCE: 12804-011W01
; CURRENT APPLICATION NUMBER: PCT/US03/14382A
; CURRENT FILING DATE: 2003-08-20
; PRIOR APPLICATION NUMBER: US 60/378,711
; PRIOR FILING DATE: 2002-05-07
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 432
; LENGTH: 4684
; TYPE: prf
; ORGANISM: Homo sapiens
PCT-US03-14382A-432

```

Query Match	10.1%;	Score 257.5;	DB 1;	Length 4684;
Best Local Similarity	20.6%;	Pred. No. 0.073;		
Matches 117;	Conservative 125;	Mismatches 196;	Indels 129;	Gaps 21;

OY	40	CGATGVPTRSSNRANS-----QAEQGEQPKYL-DSEPRKARREVEAYKKIVGSEVAST	93
Dd	1763	EAEROLIERMQLKANEALRLRLQAEIEVLQOKSLAQAEAKKOEAEAREARR-RGKAEEQAV	1822
OY	94	KKRHTITVALYNEUNINIKNEYLINKIVESTESOL-----QILMM-----ESSKVD	141
Dd	1822	RQR-----ELAEQLEKQRQLAEQTAQRRLAEBELLRLRAETEGEQQRQLLEE	1872
OY	142	VSKEPEKSSSSS-----SSDSSTKPEASDTAKPKP-----	172
Dd	1873	LARLOREAAATQQRQELAEELAKVRAEMEVLLASKAKAAEESRSTSEKSKQRLEAEAGR	1932
OY	173	-TEPEGEKVAEAKKVEEAKKAKADQEEPRNPVPTITTYTLELEIA-----ESV	221
Dd	1933	FRELAEFEAARLRALAEAEAKRORLEEBEPAARQRAEAEVLEBKLLAIGATRLKTEA	1992
OY	222	VKKAELELVKRYA-----NEPRQEQIKQAEAEVESQAEATRLKKIKITDREI-AE	272
Dd	1993	LKEKEAEENERTRLRAEDFAFORRRRLTEQEAQHKADIERLRLQTRVASELEERQGLVED	2052
OY	273	EAKRRAADAK-----GKPR-----GKPRGVPGELATPCKNDKASDSS	314
Dd	2053	TLRPRQVEEBITLAKKAFEPKAAAGKALELELGTGRNABEDTLRSKQAELEAARQRL	2112

```

QY 315 VGEETLPSPSXPEKVKAEKATKVEAAKKKDDOEEDBRNVPNTVYTLTELEISSEVE 374
Db 2113 AAEE-----ERRRREAEERQ--KSLAAEEBAARQKALBEVRLAKVBAEASL 216
QY 375 VKAE-----LELVKEEA-KEPRNEEK----VKQAKAEVSESKAEATRLKIKTDRRK 422
Db 2162 RERAEQESARQLOTLAEQAOKRLOAEEEKAHAFVQOKEQHO----QTLQOEQSVLDBRL 221
QY 423 AEEBAKRRKAAEDDKVEKPAEQP-----QAPAPKAEKPAAPKPEPAPQOPKAEKPAD 476
Db 2218 GEAPAPAPPAEAEAEAPARVQAEAREAAQSRQVEAEERLKQSAEEOQAAPAOQAAPAEK-LR 2276
QY 477 QQAEEDEVARRSEEEYNRLTQOQOPKTE 503
Db 2277 KEAEQEAARRQAEOQALRQQAADAE 2303

```

Search completed: November 21, 2003, 13:41:38  
Job time : 22.9667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: November 21, 2003, 13:32:02 ; Search time 246.996 Seconds  
(without alignments)  
1882.497 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFASKSRKRVHYSIKRFSGV.....NRLTQQCPKTEKPAQFSTP 511

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents\_AA Main:\*

1: /cgn2\_6/ptodata/1/paa/US06 COMB pep:\*

2: /cgn2\_6/ptodata/1/paa/US06 COMB pep:\*

3: /cgn2\_6/ptodata/1/paa/US07 COMB pep:\*

4: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

5: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

6: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

7: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

8: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

9: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

10: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

11: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

12: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

13: /cgn2\_6/ptodata/1/paa/US08 COMB pep:\*

14: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

15: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

16: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

17: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

18: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

19: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

20: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

21: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

22: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

23: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

24: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

25: /cgn2\_6/ptodata/1/paa/US09 COMB pep:\*

26: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

27: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

28: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

29: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

30: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

31: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

32: /cgn2\_6/ptodata/1/paa/US10 COMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2537	100.0	511	US-09-298-523B-13	Sequence 13, Appl

2	2537	100.0	511	16	US-09-298-523C-13	Sequence 13, Appl
3	2537	100.0	511	21	US-09-748-875-13	Sequence 13, Appl
4	2537	100.0	511	29	US-10-341-201-13	Sequence 13, Appl
5	2526	99.6	513	16	US-09-298-523B-12	Sequence 12, Appl
6	2526	99.6	513	16	US-09-298-523C-12	Sequence 12, Appl
7	2526	99.6	513	21	US-09-748-875-12	Sequence 12, Appl
8	2526	99.6	513	29	US-10-341-201-12	Sequence 12, Appl
9	2516	99.2	693	22	US-09-769-787-185	Sequence 15, App
10	2516	99.2	693	29	US-10-333-120A-10	Sequence 10, Appl
11	2350.5	92.6	655	14	US-09-056-019-2	Sequence 2, Appl
12	2350.5	92.6	655	14	US-09-056-019B-2	Sequence 2, Appl
13	2350.5	92.6	655	16	US-09-287-788-2	Sequence 2, Appl
14	2252.5	88.8	460	14	US-09-056-019-38	Sequence 38, Appl
15	2252.5	88.8	460	14	US-09-056-019B-38	Sequence 38, Appl
16	2252.5	88.8	460	16	US-09-287-788-38	Sequence 38, Appl
17	2235.5	88.1	459	14	US-09-056-019-39	Sequence 39, Appl
18	2235.5	88.1	459	14	US-09-056-019B-39	Sequence 39, Appl
19	2235.5	88.1	459	16	US-09-287-788-39	Sequence 39, Appl
20	2194.5	86.5	446	28	US-10-254-995-9	Sequence 9, Appl
21	2177.5	85.8	446	28	US-10-254-995-6	Sequence 6, Appl
22	2105	83.0	451	16	US-09-298-523B-67	Sequence 67, Appl
23	2105	83.0	451	16	US-09-298-523C-67	Sequence 67, Appl
24	2105	83.0	451	21	US-09-748-875-67	Sequence 67, Appl
25	2105	83.0	451	29	US-10-341-201-67	Sequence 24, Appl
26	2101.5	82.8	428	14	US-09-056-019-24	Sequence 24, Appl
27	2101.5	82.8	428	14	US-09-056-019B-24	Sequence 24, Appl
28	2101.5	82.8	428	16	US-09-287-788-24	Sequence 24, Appl
29	1979.5	78.0	406	14	US-09-056-019-1	Sequence 1, Appl
30	1979.5	78.0	406	14	US-09-056-019B-1	Sequence 1, Appl
31	1979.5	78.0	406	16	US-09-287-788-1	Sequence 1, Appl
32	1770.5	69.8	487	29	US-10-333-120A-9	Sequence 9, Appl
33	1761.5	69.4	487	16	US-09-298-523B-9	Sequence 9, Appl
34	1761.5	69.4	487	16	US-09-298-523C-9	Sequence 9, Appl
35	1761.5	69.4	487	21	US-09-748-875-9	Sequence 9, Appl
36	1761.5	69.4	487	29	US-10-341-201-9	Sequence 9, Appl
37	1745.5	68.8	487	16	US-09-298-523B-66	Sequence 66, Appl
38	1745.5	68.8	487	16	US-09-298-523C-66	Sequence 66, Appl
39	1745.5	68.8	487	21	US-09-748-875-66	Sequence 66, Appl
40	1745.5	68.8	487	29	US-10-341-201-66	Sequence 66, Appl
41	1730	68.2	701	28	US-10-282-122A-74228	Sequence 74228, A
42	1726	68.0	701	29	US-10-333-120A-7	Sequence 7, Appl
43	1718	67.7	581	16	US-09-298-523B-56	Sequence 56, Appl
44	1718	67.7	581	16	US-09-298-523C-56	Sequence 56, Appl
45	1718	67.7	581	21	US-09-748-875-56	Sequence 56, Appl

## ALIGNMENTS

RESULT 1

US-09-298-523B-13

Sequence 13, Application US/09298523B

GENERAL INFORMATION:

APPLICANT: BRILES et al.

TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS

FILE REFERENCE: 454312-3140

CURRENT APPLICATION NUMBER: US/09/298, 523B

CURRENT FILING DATE: 1999-04-23

NUMBER OF SEQ ID NOS: 78

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 13

LENGTH: 511

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-298-523B-13

Query Match 100.0%; Score 2537; DB 16; Length 511;

Best Local Similarity 100.0%; Pred. No. 1.6e-129;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MFASKSRKRVHYSIKRFSGVAVVSLVNGSVVHATENGATOVPTSSRNASQAEQ 60

```
Db 1 MFASKSERKRVHSIRKFSVGSVVAASLVMGSVVHATENEGATQVPTSSNRANESQAEQG 60
Qy 61 EOPKLDSEBPDARKEVEEYVKIIGESYAKSTKRHTITVALVNELNNIKNEYLINKIVE 120
Db 61 EOPKLDSEBPDARKEVEEYVKIIGESYAKSTKRHTITVALVNELNNIKNEYLINKIVE 120
Qy 121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSTKPEASDTAKPNKPTBGEKVA 180
Db 121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSTKPEASDTAKPNKPTBGEKVA 180
Qy 181 EAKKVEBAKAKAKQKEDRNNYPITTYKTLELEIAESDVEYKAELELVKYNANEPDE 240
Db 181 EAKKVEBAKAKAKQKEDRNNYPITTYKTLELEIAESDVEYKAELELVKYNANEPDE 240
Qy 241 OKIKQAEAEVSKQAEATRLKIKTDREBAEBAERADAKQOGPKGRAGVPGELAT 300
Db 241 OKIKQAEAEVSKQAEATRLKIKTDREBAEBAERADAKQOGPKGRAGVPGELAT 300
Qy 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEBAKKAEDQKEDRNNYPINT 360
Db 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEBAKKAEDQKEDRNNYPINT 360
Qy 361 YKTLELEIAESDVEYKAELELVKEBAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Db 361 YKTLELEIAESDVEYKAELELVKEBAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Qy 421 KKAEEBAKKAEEBKVKEKPAEQOPAPAPAEKPAAPKPNBAPAEQPKAEKPADQOAE 480
Db 421 KKAEEBAKKAEEBKVKEKPAEQOPAPAPAEKPAAPKPNBAPAEQPKAEKPADQOAE 480
Qy 481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511
Db 481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511

RESULT 2
US-09-298-523C-13
; Sequence 13, Application US/09298523C
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523C
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523C-13

Query Match 100.0%; Score 2537; DB 16; Length 511;
Best Local Similarity 100.0%; Pred. No. 1,66-129;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 241 OKIKQAEAEVSKQAEATRLKIKTDREBAEBAERADAKQOGPKGRAGVPGELAT 300
Db 241 OKIKQAEAEVSKQAEATRLKIKTDREBAEBAERADAKQOGPKGRAGVPGELAT 300
Qy 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEBAKKAEDQKEDRNNYPINT 360
Db 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEBAKKAEDQKEDRNNYPINT 360
Qy 361 YKTLELEIAESDVEYKAELELVKEBAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Db 361 YKTLELEIAESDVEYKAELELVKEBAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Qy 421 KKAEEBAKKAEEBKVKEKPAEQOPAPAPAEKPAAPKPNBAPAEQPKAEKPADQOAE 480
Db 421 KKAEEBAKKAEEBKVKEKPAEQOPAPAPAEKPAAPKPNBAPAEQPKAEKPADQOAE 480
Qy 481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511
Db 481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511

RESULT 3
US-09-748-875-13
; Sequence 13, Application US/09748875
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-13

Query Match 100.0%; Score 2537; DB 21; Length 511;
Best Local Similarity 100.0%; Pred. No. 1,66-129;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
QY 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511  
DB 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511

## RESULT 4

US-10-341-201-13  
; Sequence 13, Application US/10341201  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/10/341,201  
; CURRENT FILING DATE: 2003-01-13  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-341-201-13

Query Match 100.0%; Score 2537; DB 29; Length 511;  
Best Local Similarity 100.0%; Pred. No. 1,6e-129;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPASKEKRVHYSIRKFSVGASVVAASLVMSGVHATNEGATQVPTSSNRANESQAEQG 60  
DB 1 MPASKEKRVHYSIRKFSVGASVVAASLVMSGVHATNEGATQVPTSSNRANESQAEQG 60  
QY 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
DB 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
QY 121 STSSQOILMMESRSKYDEAVSKFEKDSSTSSSTKPEASDTAKPNKPTBGEKVA 180  
DB 121 STSSQOILMMESRSKYDEAVSKFEKDSSTSSSTKPEASDTAKPNKPTBGEKVA 180  
QY 181 EAKKVEBAKKAQOKEEDRRNYPITYKTLELEIAESDVVKAELELVKANKNEPRDE 240  
DB 181 EAKKVEBAKKAQOKEEDRRNYPITYKTLELEIAESDVVKAELELVKANKNEPRDE 240  
QY 241 OKIQAEEAVESKQAEATRLKKITDRBEAEEBAKRAADAKEOGKPKGRAGVPGELAT 300  
DB 241 OKIQAEEAVESKQAEATRLKKITDRBEAEEBAKRAADAKEOGKPKGRAGVPGELAT 300  
QY 301 PDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKEAKKAEDQKEEDRRNYPNT 360  
DB 301 PDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKEAKKAEDQKEEDRRNYPNT 360  
QY 361 YKTELEIAESDVVKAELELVKEAKEPEPNEKVKQAKAEVSKAEATRLKIKITDR 420  
DB 361 YKTELEIAESDVVKAELELVKEAKEPEPNEKVKQAKAEVSKAEATRLKIKITDR 420  
QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
QY 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511  
DB 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511

## RESULT 5

US-09-298-523B-12  
; Sequence 12, Application US/09298523B  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.

; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523B  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-12

Query Match 99.6%; Score 2526; DB 16; Length 513;  
Best Local Similarity 99.6%; Pred. No. 6,2e-129;  
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPASKEKRVHYSIRKFSVGASVVAASLVMSGVHATNEGATQVPTSSNRANESQAEQG 60  
DB 1 MPASKEKRVHYSIRKFSVGASVVAASLVMSGVHATNEGATQVPTSSNRANESQAEQG 60  
QY 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
DB 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
QY 121 STSSQOILMMESRSKYDEAVSKFEKDSSTSSSTKPEASDTAKPNKPTBGEKVA 180  
DB 121 STSSQOILMMESRSKYDEAVSKFEKDSSTSSSTKPEASDTAKPNKPTBGEKVA 180  
QY 181 EAKKVEBAKKAQOKEEDRRNYPITYKTLELEIAESDVVKAELELVKANKNEPRDE 240  
DB 181 EAKKVEBAKKAQOKEEDRRNYPITYKTLELEIAESDVVKAELELVKANKNEPRDE 240  
QY 241 OKIQAEEAVESKQAEATRLKKITDRBEAEEBAKRAADAKEOGKPKGRAGVPGELAT 300  
DB 241 OKIQAEEAVESKQAEATRLKKITDRBEAEEBAKRAADAKEOGKPKGRAGVPGELAT 300  
QY 301 PDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKEAKKAEDQKEEDRRNYPNT 360  
DB 301 PDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKKEAKKAEDQKEEDRRNYPNT 360  
QY 361 YKTELEIAESDVVKAELELVKEAKEPEPNEKVKQAKAEVSKAEATRLKIKITDR 420  
DB 361 YKTELEIAESDVVKAELELVKEAKEPEPNEKVKQAKAEVSKAEATRLKIKITDR 420  
QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEOPKAKEPADQOAE 480  
QY 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511  
DB 481 EDYARSEEEYNRLTQOQPPKTEKPAQPSSTP 511

## RESULT 6

US-09-298-523C-12  
; Sequence 12, Application US/09298523C  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/298,523C  
; CURRENT FILING DATE: 1999-04-23  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-298-523C-12

Query Match 99.6%; Score 2526; DB 16; Length 513;

Best Local Similarity 99.6%; Pred. No. 6.2e-129;  
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFASRERKRVHSIRKFSVGSVVVAASLVMSGVHATENEGATQVPTSSNRANESQAOEG 60  
DB 1 MFASRERKRVHSIRKFSVGSVVVAASLVMSGVHATENEGATQVPTSSNRANESQAOEG 60  
QY 61 EOPKLDSEBDKARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
DB 61 EOPKLDSEBDKARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
QY 121 STSESQLOILMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTGEKVA 180  
DB 121 STSESQLOILMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTGEKVA 180  
QY 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
DB 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
QY 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
DB 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
QY 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360  
DB 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360  
QY 361 YKTTLEIAESDVEYKKAELVKEBAKEPRNEEKVKQAKAEVSKKAATRLKIKTDR 420  
DB 361 YKTTLEIAESDVEYKKAELVKEBAKEPRNEEKVKQAKAEVSKKAATRLKIKTDR 420  
QY 421 KKAEEBAKKAEDKVEKPAEQOPAPAPAKAPAPKAPENPAEQPKAKPADQOAE 480  
DB 421 KKAEEBAKKAEDKVEKPAEQOPAPAPAKAPAPKAPENPAEQPKAKPADQOAE 480  
QY 481 EDYARSEEEYRNLTOOQPPKTEKPAQSTP 511  
DB 481 EDYARSEEEYRNLTOOQPPKTEKPAQSTP 511

RESULT 7  
US-09-748-875-12  
; Sequence 12, Application US/09748875  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/09/748,875  
; PRIOR FILING DATE: 2000-12-26  
; PRIOR APPLICATION NUMBER: 09/298,523  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-748-875-12

Query Match 99.6%; Score 2526; DB 21; Length 513;  
Best Local Similarity 99.6%; Pred. No. 6.2e-129;  
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 121 STSESQLOILMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTGEKVA 180  
QY 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
DB 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
QY 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
DB 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
QY 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360  
DB 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360  
QY 361 YKTTLEIAESDVEYKKAELVKEBAKEPRNEEKVKQAKAEVSKKAATRLKIKTDR 420  
DB 361 YKTTLEIAESDVEYKKAELVKEBAKEPRNEEKVKQAKAEVSKKAATRLKIKTDR 420  
QY 421 KKAEEBAKKAEDKVEKPAEQOPAPAPAKAPAPKAPENPAEQPKAKPADQOAE 480  
DB 421 KKAEEBAKKAEDKVEKPAEQOPAPAPAKAPAPKAPENPAEQPKAKPADQOAE 480  
QY 481 EDYARSEEEYRNLTOOQPPKTEKPAQSTP 511  
DB 481 EDYARSEEEYRNLTOOQPPKTEKPAQSTP 511

RESULT 8  
US-10-341-201-12  
; Sequence 12, Application US/10341201  
; GENERAL INFORMATION:  
; APPLICANT: BRILES et al.  
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
; FILE REFERENCE: 454312-3140  
; CURRENT APPLICATION NUMBER: US/10/341,201  
; PRIOR FILING DATE: 2003-01-13  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-341-201-12

Query Match 99.6%; Score 2526; DB 29; Length 513;  
Best Local Similarity 99.6%; Pred. No. 6.2e-129;  
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFASRERKRVHSIRKFSVGSVVVAASLVMSGVHATENEGATQVPTSSNRANESQAOEG 60  
DB 1 MFASRERKRVHSIRKFSVGSVVVAASLVMSGVHATENEGATQVPTSSNRANESQAOEG 60  
QY 61 EOPKLDSEBDKARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
DB 61 EOPKLDSEBDKARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNNIKNEYLINKIVE 120  
QY 121 STSESQLOILMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTGEKVA 180  
DB 121 STSESQLOILMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTGEKVA 180  
QY 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
DB 181 EAKKVEBAKAKADQKEDRRNYPTITYKTTLEIAESDVEYKKAELVVKANEPDE 240  
QY 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
DB 241 OKIQAEEVSKQAEATRLKKTDRBEAEBAERADAKGPKGRARAGVPGELAT 300  
QY 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360  
DB 301 PDKENDAKSSDSSVGEETLPSPLKPEKVAEAKVEAKKAEDQKEDRRNYPTNT 360

```
Qy 361 YKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKTDR 420
| | | | |
Db 361 YKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKTDR 420
Qy 421 KKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQOAE 480
| | | | |
Db 421 KKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQOAE 480
Qy 481 EDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 511
| | | | |
Db 481 EDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 511

RESULT 9
US-09-769-787-185
; Sequence 185, Application US/09769787
; GENERAL INFORMATION:
; APPLICANT: Microbial Techniques Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P2129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 185
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-185

Query Match 99.2%; Score 2516; DB 22; Length 693;
Best Local Similarity 99.6%; Pred. No. 3.2e-128;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 MFASKSRKRVHSIRKFSVG-ASVVVASLYMGSVVHATENEGATQVPTSSNRANESQAEQ 59
| | | | |
Db 1 MFASKSRKRVHSIRKFSVGASVVVASLYMGSVVHATENEGATQVPTSSNRANESQAEQ 60
Qy 60 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 119
| | | | |
Db 60 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 119
Qy 61 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 120
| | | | |
Db 61 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 120
Qy 120 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 179
| | | | |
Db 120 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
Qy 121 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
| | | | |
Db 121 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
Qy 180 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 238
| | | | |
Db 180 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
Qy 181 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
| | | | |
Db 181 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
Qy 239 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 298
| | | | |
Db 239 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
Qy 241 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
| | | | |
Db 241 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
Qy 299 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 358
| | | | |
Db 299 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
Qy 301 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
| | | | |
Db 301 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
Qy 359 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 418
| | | | |
Db 359 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
Qy 361 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
| | | | |
Db 361 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
Qy 419 DRKKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQO 478
| | | | |
Db 419 DRKKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQO 480
Qy 421 DRKKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQO 480
| | | | |
Db 421 DRKKAEEAKRKAEDVKKEPAEQOPAPAPAKKAPAPKPPENPAEQPKAEKPADQO 480
```

```
Qy 479 AEDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 511
| | | | |
Db 481 AEDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 513

RESULT 10
US-10-333-120A-10
; Sequence 10, Application US/10333120A
; GENERAL INFORMATION:
; APPLICANT: Danulczyk, Robert
; APPLICANT: Iannelli, Francesco
; APPLICANT: Sjoelund, Anders
; APPLICANT: Pozzi, Gianni
; APPLICANT: Björck, Lars
; TITLE OF INVENTION: PROTEIN
; FILE REFERENCE: 13519PCTUS
; CURRENT APPLICATION NUMBER: US/10/333,120A.
; PRIOR APPLICATION NUMBER: PCT/EP01/08409
; PRIOR FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: SE 0002728-4
; PRIOR FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-333-120A-10

Query Match 99.2%; Score 2516; DB 29; Length 693;
Best Local Similarity 99.6%; Pred. No. 3.2e-128;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
```

```
Qy 1 MFASKSRKRVHSIRKFSVG-ASVVVASLYMGSVVHATENEGATQVPTSSNRANESQAEQ 59
| | | | |
Db 1 MFASKSRKRVHSIRKFSVGASVVVASLYMGSVVHATENEGATQVPTSSNRANESQAEQ 60
Qy 60 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 119
| | | | |
Db 60 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 119
Qy 61 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 120
| | | | |
Db 61 GEOPKULDSEDRDARKKEVEEYVKKI VGESYAKSTKKHTTTVALVNLNINIKNEYLNKIY 120
Qy 120 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 179
| | | | |
Db 120 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
Qy 121 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
| | | | |
Db 121 ESTSESOLQILMMESRKVDVAVSKFEKDSSSSSSDSTKPEASDPAKPKPTEPGEKV 180
Qy 180 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 238
| | | | |
Db 180 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
Qy 181 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
| | | | |
Db 181 AEAKKKVEEA-KKAKDQKEEDRRNYPITTYKLELEIAESDVEVKAELELVKVKANEP 240
Qy 239 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 298
| | | | |
Db 239 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
Qy 241 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
| | | | |
Db 241 DEOKIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAKEQKPKGAKGVPGE 300
Qy 299 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 358
| | | | |
Db 299 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
Qy 301 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
| | | | |
Db 301 ATPDKKENDAKSSDSSGEEITLPSPLKPEKKVAEAKKYVEAKKAKEDQKEEDRRNYP 360
Qy 359 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 418
| | | | |
Db 359 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
Qy 361 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
| | | | |
Db 361 NTYKLELEIAESDVEVKAELELVKEEAKPRNEEKVQAKAEVSSKKAETRLKIKT 420
Qy 479 AEDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 511
| | | | |
Db 481 AEDVARSSEEEYNRLTQOOPPKTEKPAOPSTP 513

RESULT 11
US-09-056-019-2
```



Sequence 2, Application US/09056019A  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzmann, Theresa  
APPLICANT: Maure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017  
CURRENT APPLICATION NUMBER: US/09/056, 019A  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019-2

Query Match 92.6%; Score 2350.5; DB 14; Length 655;  
Best Local Similarity 99.8%; Pred. No. 3e-119;  
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATOVPTSSNRANESQAEQGEOPKLDSEPRDKARKEVEYVKIIGESYAKSTKRRH 97  
DB 1 ENEGATOVPTSSNRANESQAEQGEOPKLDSEPRDKARKEVEYVKIIGESYAKSTKRRH 60  
QY 98 TITVALVNLNNIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSDS 157  
DB 61 TITVALVNLNNIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSDS 120  
QY 158 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEEDRRNYPTTYKTLELEIA 216  
DB 121 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEEDRRNYPTTYKTLELEIA 180  
QY 217 ESDVEVKKAELVYKVA NEPRDEQIKQA EAVESKQAEATRLKIKITDREAEAEBAKR 276  
DB 181 ESDVEVKKAELVYKVA NEPRDEQIKQA EAVESKQAEATRLKIKITDREAEAEBAKR 240  
QY 277 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 336  
DB 241 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 300  
QY 337 KVEEAKKKAEDQKEDRRNYPTNYKTLELEIAESDVYKKAELVYKEAEKEPRNEKV 396  
DB 301 KVEEAKKKAEDQKEDRRNYPTNYKTLELEIAESDVYKKAELVYKEAEKEPRNEKV 360  
QY 397 KOAKAEVSKKAETRLKIKITDRKKAEEBAKRAAEEDVYKEKPAEQPOPAAPAKAEKP 456  
DB 361 KOAKAEVSKKAETRLKIKITDRKKAEEBAKRAAEEDVYKEKPAEQPOPAAPAKAEKP 420  
QY 457 AAPKPEPAPQPAEKPADQAEEDVYARSEEEYNRLTQOQPPKTEKPAQPSPT 511  
DB 421 AAPKPEPAPQPAEKPADQAEEDVYARSEEEYNRLTQOQPPKTEKPAQPSPT 475

RESULT 12  
US-09-056-019B-2  
Sequence 2, Application US/09056019B  
GENERAL INFORMATION:  
APPLICANT: Tuomanen, Elaine I  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Maure, H. R.  
APPLICANT: Johnson, Leslie S.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 5853-2  
CURRENT APPLICATION NUMBER: US/09/056, 019B  
CURRENT FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 40  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-056-019B-2

Query Match 92.6%; Score 2350.5; DB 14; Length 655;  
Best Local Similarity 99.8%; Pred. No. 3e-119;  
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATOVPTSSNRANESQAEQGEOPKLDSEPRDKARKEVEYVKIIGESYAKSTKRRH 97  
DB 1 ENEGATOVPTSSNRANESQAEQGEOPKLDSEPRDKARKEVEYVKIIGESYAKSTKRRH 60  
QY 98 TITVALVNLNNIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSDS 157  
DB 61 TITVALVNLNNIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSDS 120  
QY 158 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEEDRRNYPTTYKTLELEIA 216  
DB 121 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEEDRRNYPTTYKTLELEIA 180  
QY 217 ESDVEVKKAELVYKVA NEPRDEQIKQA EAVESKQAEATRLKIKITDREAEAEBAKR 276  
DB 181 ESDVEVKKAELVYKVA NEPRDEQIKQA EAVESKQAEATRLKIKITDREAEAEBAKR 240  
QY 277 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 336  
DB 241 RADAKEQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 300  
QY 337 KVEEAKKKAEDQKEDRRNYPTNYKTLELEIAESDVYKKAELVYKEAEKEPRNEKV 396  
DB 301 KVEEAKKKAEDQKEDRRNYPTNYKTLELEIAESDVYKKAELVYKEAEKEPRNEKV 360  
QY 397 KOAKAEVSKKAETRLKIKITDRKKAEEBAKRAAEEDVYKEKPAEQPOPAAPAKAEKP 456  
DB 361 KOAKAEVSKKAETRLKIKITDRKKAEEBAKRAAEEDVYKEKPAEQPOPAAPAKAEKP 420  
QY 457 AAPKPEPAPQPAEKPADQAEEDVYARSEEEYNRLTQOQPPKTEKPAQPSPT 511  
DB 421 AAPKPEPAPQPAEKPADQAEEDVYARSEEEYNRLTQOQPPKTEKPAQPSPT 475

RESULT 13  
US-09-287-788-2  
Sequence 2, Application US/09287788B  
GENERAL INFORMATION:  
APPLICANT: Maure, H. R.  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
FILE REFERENCE: 1340-1-017N  
CURRENT APPLICATION NUMBER: US/09/287, 788B  
CURRENT FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: 60/080, 878  
PRIOR FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-287-788-2

Query Match 92.6%; Score 2350.5; DB 16; Length 655;  
Best Local Similarity 99.8%; Pred. No. 3e-119;  
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 38 ENEGATOVPTSSNRANESQAEQGEOPKLDSEPRDKARKEVEYVKIIGESYAKSTKRRH 97



Db 1 ENEGATVPTSSNRANESQAEQEQPKLDSERDVARKEVEYKKIVGESYAKSTKKRH 60  
Qy 98 TTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVSKFEKSSSSSSDS 157  
Db 61 TTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVSKFEKSSSSSSDS 120  
Qy 158 STKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRNPTTYKTLELEIA 216  
Db 121 STKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRNPTTYKTLELEIA 180  
Qy 217 ESDVEVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKKITDSEBAEBAKR 276  
Db 181 ESDVEVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKKITDSEBAEBAKR 240  
Qy 217 RADAEQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPSPLKEKKVAEAK 336  
Db 241 RADAEQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPSPLKEKKVAEAK 300  
Qy 337 KVEEAKKAEADQKEEDRRNPTTYKTLELEIAESDVVKAELELVKEBAKEPRNEEKV 396  
Db 301 KVEEAKKAEADQKEEDRRNPTTYKTLELEIAESDVVKAELELVKEBAKEPRNEEKV 360  
Qy 397 KOAKAEVSKKAEATRLKIKTRKKAEBAKRAKAEEDKYKEPAEQPPAPAKAEKP 456  
Db 361 KOAKAEVSKKAEATRLKIKTRKKAEBAKRAKAEEDKYKEPAEQPPAPAKAEKP 420  
Qy 457 APAPKPPAPAKAEADQAEEDVARSSEENRLLTQOOPKTEKPAQSP 511  
Db 421 APAPKPPAPAKAEADQAEEDVARSSEENRLLTQOOPKTEKPAQSP 475

RESULT 14  
US-09-056-019-38  
; Sequence 38, Application US/09056019A  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wizemann, Theresa  
; APPLICANT: Maure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Koenig, Scott  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; FILE REFERENCE: 1340-1-017  
; CURRENT APPLICATION NUMBER: US/09/056,019A  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019-38

Query Match 88.8%; Score 2252.5; DB 14; Length 460;  
Best Local Similarity 99.6%; Pred. No. 4.1e-114;  
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 24 VVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDVARKEVEYKK 83  
Db 2 IVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDVARKEVEYKK 61  
Qy 84 IVGESYAKSTKKRTTTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVS 143  
Db 62 IVGESYAKSTKKRTTTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVS 121  
Qy 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRN 202  
Db 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRN 181  
Qy 203 YPTITTYKTLELEIAESDVVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKK 262  
Db 182 YPTITTYKTLELEIAESDVVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKK 241

Qy 263 IKTDREAEBAKRRADAKQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPS 322  
Db 242 IKTDREAEBAKRRADAKQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPS 301  
Qy 323 PSLKPEKKVAEAEKVEBAKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELEL 382  
Db 302 PSLKPEKKVAEAEKVEBAKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELEL 361  
Qy 383 VKEBAKEPRNEEKVKOAKAEVSKKAEATRLKIKTRKKAEBAKRAAEEDKYKEPA 442  
Db 362 VKEBAKEPRNEEKVKOAKAEVSKKAEATRLKIKTRKKAEBAKRAAEEDKYKEPA 421  
Qy 443 EQPPAPAKAEKPAPAKPPENPAEQPAEKPADQAE 481  
Db 422 EQPPAPAKAEKPAPAKPPENPAEQPAEKPADQAE 460

RESULT 15  
US-09-056-019B-38  
; Sequence 38, Application US/09056019B  
; GENERAL INFORMATION:  
; APPLICANT: Tuomanen, Elaine I  
; APPLICANT: Wizemann, Theresa M.  
; APPLICANT: Maure, H. R.  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Koenig, Scott  
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL  
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED  
; FILE REFERENCE: 5853-2  
; CURRENT APPLICATION NUMBER: US/09/056,019B  
; CURRENT FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-056-019B-38

Query Match 88.8%; Score 2252.5; DB 14; Length 460;  
Best Local Similarity 99.6%; Pred. No. 4.1e-114;  
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 24 VVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDVARKEVEYKK 83  
Db 2 IVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDVARKEVEYKK 61  
Qy 84 IVGESYAKSTKKRTTTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVS 143  
Db 62 IVGESYAKSTKKRTTTTVALVNELNINKIEYLINKIVESTSESOLIMMESRSKYDEAVS 121  
Qy 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRN 202  
Db 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTBEGEKVAEAKKVEEA-KKAKDQKEEDRRN 181  
Qy 203 YPTITTYKTLELEIAESDVVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKK 262  
Db 182 YPTITTYKTLELEIAESDVVKAELELVKVAKANEPDEOKIKOAEAVESKQAEATRLKK 241  
Qy 263 IKTDREAEBAKRRADAKQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPS 322  
Db 242 IKTDREAEBAKRRADAKQKPKGRARAGVPELATPDKKENDAKSSDSSVGEETLPS 301  
Qy 323 PSLKPEKKVAEAEKVEBAKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELEL 382  
Db 302 PSLKPEKKVAEAEKVEBAKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELEL 361  
Qy 383 VKEBAKEPRNEEKVKOAKAEVSKKAEATRLKIKTRKKAEBAKRAAEEDKYKEPA 442  
Db 362 VKEBAKEPRNEEKVKOAKAEVSKKAEATRLKIKTRKKAEBAKRAAEEDKYKEPA 421

Mon Nov 24 10:00:24 2003

us-09-298-523c-13.رام

Page 8

QY 443 EQPPAPAPKAEPPAPAPKPPNPAPQPKAEKPPADQAAEE 481  
Dp 422 EQPPAPAPKAEKPPAPAPKPPNPAPQPKAEKPPADQAAEE 460

Search completed: November 21, 2003, 13:40:57  
Job time : 247.996 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: November 21, 2003, 13:29:32 ; Search time 20.7062 Seconds  
(without alignments)  
2373.306 Million cell updates/sec

Title: US-09-298-523C-13  
Perfect score: 2537  
Sequence: 1 MFASKSRKRVHYSIRKFSVG.....NRLTQQQPPKTEKPAQSTP 511

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR 76:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	2 H95255	choleline binding pr
2	1730	68.2	701	2 H98120	choleline binding pr
3	491	19.4	619	2 A41971	surface protein ps
4	491	19.4	619	2 A97887	surface protein ps
5	397.5	15.7	744	2 F95013	pneumococcal surfa
6	389.5	15.4	1110	2 I51116	NR-180 - sea lampr
7	363	14.3	1020	1 QPHUH	neurofilament trip
8	361	14.2	1701	2 T09127	probable erythrocy
9	351.5	13.9	1164	1 FCSOAG	IGA Fc receptor pr
10	346.5	13.7	1134	2 A60234	IGA Fc receptor pr
11	345	13.6	606	2 A43477	neurofilament trip
12	337.5	13.3	1052	2 A44937	kineoclast-associ
13	329.5	13.0	3488	2 T34418	hypothetical prote
14	328	12.9	1130	2 S51364	sperm tail-specifi
15	328	12.9	6642	2 T29757	protein UNC-89 - C
16	322	12.7	1072	1 A37221	neurofilament trip
17	318	12.5	854	2 S02003	neurofilament trip
18	315.5	12.4	706	2 A45980	junctional sarcopl
19	315	12.4	522	2 C96608	hypothetical prote
20	313.5	12.4	699	2 E84565	hypothetical prote
21	311	12.3	539	2 A28549	M24 protein precur
22	310	12.2	1087	1 QFMH	neurofilament trip
23	308	12.1	5327	2 T13564	microtubule-associ
24	300	11.8	771	1 A33430	h-caldesmon - chic
25	293	11.5	1526	2 A45605	mature-parasite-in
26	292.5	11.5	858	2 S15762	neurofilament trip
27	292.5	11.5	1815	2 C81169	IGA-specific metal
28	292	11.5	1094	2 S49313	protein kinase - s
29	290	11.4	729	2 S68191	triadin - human

30	288	11.4	1128	2 T30296	R27-2 protein - Tr
31	285.5	11.3	501	2 A44643	M protein precursor
32	285.5	11.3	1773	2 A81937	IGA-specific metal
33	284	11.2	644	2 S55395	neurofilament prot
34	283	11.2	916	2 A27864	neurofilament trip
35	282.5	11.1	849	2 S00030	neurofilament trip
36	279	11.0	667	2 A40713	cylicin I - bovine
37	278	11.0	488	2 I46014	cylicin II - bovin
38	275	10.8	488	2 A71623	probable secreted
39	274.5	10.8	332	2 B43427	neurofilament prot
40	274	10.8	2464	1 QRMSP1	microtubule-associ
41	270.5	10.7	598	2 B40713	cylicin I - human
42	269.5	10.6	532	2 S54871	M protein - Strept
43	268.5	10.6	798	2 I50479	neurofilament medi
44	268.5	10.6	2364	2 A56577	microtubule-associ
45	266	10.5	483	2 A26297	M6 protein - Strept

## ALIGNMENTS

RESULT 1					
H95255					
choleline binding protein A [imported] - Streptococcus pneumoniae (strain TIGR4)					
C/Species: Streptococcus pneumoniae					
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001					
C/Accession: H95255					
R/tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid					
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,					
nson, T.; Hickey, E.K.; Holt, I.E.					
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,					
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.					
A/Reference number: A95000; MUID:21357209; PMID:11463916					
A/Accession: H95255					
A/Status: preliminary					
A/Molecule type: DNA					
A/Residues: 1-693 <NR>					
A/Cross-references: GB:AE005672; PIDN:AAK76241.1; PID:G14973701; GSPDB:GN00164; TIGR:SP4					
A/Experimental source: strain TIGR4					
C/Genetics:					
A/Gene: SP2190					
Query Match					
Best Local Similarity 99.2%; Score 2516; DB 2; Length 693;					
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;					
QY	1	MFASKSRKRVHYSIRKFSVG-ASVVVASLVGSVVHATNEGATQVPTSSNRANESQAEQ	59		
DB	1	MFASKSRKRVHYSIRKFSVGASVVVASLVGSVVHATNEGATQVPTSSNRANESQAEQ	60		
QY	60	GEOPKLDSEDRARKEVEEYVKKIVGESYAKSTKRRHTTVALVNLNNIKNYEYLNKIV	119		
DB	61	GEOPKLDSEDRARKEVEEYVKKIVGESYAKSTKRRHTTVALVNLNNIKNYEYLNKIV	120		
QY	120	ESHSSESOLOLIMHESRSKVDKAVSKPEKDSSSSSSSSTKPEASDPKANKPKPEBEKV	179		
DB	121	ESHSSESOLOLIMHESRSKVDKAVSKPEKDSSSSSSSSTKPEASDPKANKPKPEBEKV	180		
QY	180	AEKKKVEEA-KKAKOKEEDRRNYPITYKTLELEAEEDVEKKAELVKKANEP	238		
DB	181	AEKKKVEEA-KKAKOKEEDRRNYPITYKTLELEAEEDVEKKAELVKKANEP	240		
QY	239	DEOKIQAEAEVSKOAEATRLKIKITDRBEAEAEARRADAKOQPKGRANKGVDEL	298		
DB	241	DEOKIQAEAEVSKOAEATRLKIKITDRBEAEAEARRADAKOQPKGRANKGVDEL	300		
QY	299	ATDDKKENDAKSSDSSVGERTLPSPSLKPEKKAVAEAKKVEAKKAADQKEDRRNYP	358		
DB	301	ATDDKKENDAKSSDSSVGERTLPSPSLKPEKKAVAEAKKVEAKKAADQKEDRRNYP	360		
QY	359	NTKTELELEAEEDVEKKAELVKEAEKPEENEEVKOAKAEVSKKAEATRLKIKT	418		

Db 361 NTYKTLLEIAESDVVEVKAELELVKEBAKEPRNEKVKAQAEVSKAEATRLKIKT 420  
Qy 419 DRKAAEEAKKAAAEEDVYKEKPAEQOPAPAPKPAKPAKPPENPAEQKAEKPADQ 478  
Db 421 DRKAAEEAKKAAAEEDVYKEKPAEQOPAPAPKPAKPAKPPENPAEQKAEKPADQ 480  
Qy 479 AEDYARSESEYNNLTQOOPKTEKPAQPSPT 511  
Db 481 AEDYARSESEYNNLTQOOPKTEKPAQPSPT 513

## RESULT 2

H98120  
C:Species: Streptococcus pneumoniae (strain R6)  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: H98120  
R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, T.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: H98120  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-701 <KOR>  
A:Cross-references: GB:AE07317, PIDN:AAU00797.1; PID:915459699; GSPDB:GN00174  
C:Genetics:  
A:Gene: pspc

Query Match 68.2%; Score 1730; DB 2; Length 701;  
Best Local Similarity 70.7%; Pred. No. 1,66-59;  
Matches 371; Conservative 49; Mismatches 47; Indels 58; Gaps 10;

Qy 1 MFASKSEKRVHYSRKFSVG-ASVVASLVWGVSVAHATENGATOVPTSSRAESQAEQ 59  
Db 1 MFASKSEKRVHYSRKFSVG-ASVVASLVWGVSVAHATENGATOVPTSSRAESQAEQ 54  
Qy 60 GEOPKLDSEDRKARKE-VBEYVKIVGESYAKSTKKRHITVALVNLNINKEVYANK- 117  
Db 55 -----TEHRKAAQVVDYEIEKMLRE--IQDRRHGTQVNLNITKLSAIIKTYREL 104  
Qy 118 -IVSTSESQIQLMBSRKVDVAVKFEVDSSSSSSSDSTPEASDTAKPKPTPEPG 176  
Db 105 NVLEEKSKDELPL--SEIKALDLDAFEKFKKOT-----LKKP 138  
Qy 177 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLLEIAESDVEYKAELELVKYN 235  
Db 139 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLLEIAESDVEYKAELELVKYN 198  
Qy 236 EPRDEQIKQAEVSKQAEATRLKIKITDREAESEAKRADAK-----EQQKP 286  
Db 199 ESRNEGITIKQAEVSKQAEATRLKIKITDREAESEAKRADAK-----EQQKP 258  
Qy 287 KGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSGLKREKVAEAKKVEAKKAE 346  
Db 259 KGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSGLKREKVAEAKKVEAKKAE 318  
Qy 347 DQKEEDRRNYPITTYKTLLEIAESDVEYKAELELVKEAKEPRNEKVKAQAEVSK 406  
Db 319 DQKEEDRRNYPITTYKTLLEIAESDVEYKAELELVKEAKEPRNEKVKAQAEVSK 378  
Qy 407 KAEATRLKIKITDREAESEAKKVEAKKVEAKKVEAKKVEAKKVEAKKVEAKK 466  
Db 379 KAEATRLKIKITDREAESEAKKVEAKKVEAKKVEAKKVEAKKVEAKKVEAKK 436  
Qy 467 EOPKAEKPADQAEEDYARSESEYNNLTQOOPKTEKPAQPSPT 511  
Db 437 EOPKAEKPADQAEEDYARSESEYNNLTQOOPKTEKPAQPSPT 481

## RESULT 3

A41971  
surface protein pspa precursor - Streptococcus pneumoniae  
N:Alternate names: pneumococcal surface protein A  
C:Species: Streptococcus pneumoniae  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999  
C:Accession: A41971; A60282; A33134  
R:Yoher, J.; Briles, D.E.  
J. Bacteriol. 174, 601-609, 1992  
A:Title: Structural properties and evolutionary relationships of Pspa, a surface protein  
A:Reference number: A41971; MUID:92105030; PMID:11729249  
A:Accession: A41971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-619 <YOT>  
A:Cross-references: GB:M74122, NID:9153840; PIDN:AAA2018.1; PID:9153841  
A:Note: sequence extracted from NCBI backbone (NCBIN:75635, NCBI:P:75636)  
R:Talkington, D.F.; Crimmins, L.; Voellinger, D.C.; Yoher, J.; Briles, D.E.  
Infect. Immun. 59, 1285-1289, 1991  
A:Title: A 43-kilodalton pneumococcal surface protein, Pspa: isolation, protective ability  
A:Reference number: A60282; MUID:91169596; PMID:2004810  
A:Accession: A60282  
A:Molecule type: protein  
A:Residues: 32-76 <TAL>  
A:Experimental source: strain JY2008  
C:Genetics:  
A:Gene: pspa  
A:Superfamily: cpl repeat homology  
F:1-31/Domain: signal sequence #status predicted <SIG>  
F:32-619/Product: surface protein pspa #status predicted <MAT>  
F:411-430/Domain: cpl repeat homology <CP01>  
F:431-450/Domain: cpl repeat homology <CP02>  
F:451-470/Domain: cpl repeat homology <CP03>  
F:471-490/Domain: cpl repeat homology <CP04>  
F:491-510/Domain: cpl repeat homology <CP05>  
F:511-530/Domain: cpl repeat homology <CP06>  
F:531-550/Domain: cpl repeat homology <CP07>  
F:551-570/Domain: cpl repeat homology <CP08>  
F:571-591/Domain: cpl repeat homology <CP09>  
F:592-611/Domain: cpl repeat homology <CP10>

Query Match 19.4%; Score 491; DB 2; Length 619;  
Best Local Similarity 37.3%; Pred. No. 2,7e-12;  
Matches 152; Conservative 50; Mismatches 109; Indels 96; Gaps 18;

Qy 161 PEASDTAKPKPTPEGEKVAE-AKKVVEAKKAKDQKEEDRRNYPITTYKTLLEIAESD 219  
Db 35 PVASQ-SKAEKDYDAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKKAKK 86  
Qy 220 VEYKAELELVKYNKAEPRDEQIKQAEVSKQAEATRLKIKITDREAESEAKRAD 279  
Db 87 -----KALEKVA--ASEEMDKAVALVQAVLAYQATD--KAKADAKMIDEAKR-- 134  
Qy 280 AKQGRKRGKRGV-----PGLATPDKKENDAKSSDSSVGEETLPSGLKREKVAE 335  
Db 135 -EEAKTKFNTVAAMVVPPEQLAETKKSSEAKK-----APEL--TKULEAK 181  
Qy 336 KQVEAKKKAED--QKEEDRRNYP-----INTYKTLLEIAESD----- 372  
Db 182 AKLEBAEKKAETKQKQVDAEVAPOAKIAELENVQVRLLEQLEIASESESEDAKGFRA 241  
Qy 373 -----VEYKAELELVKE-----EAKPRNEKVKAQAEVSK-----VES 405  
Db 242 PLOSKLDKAKKAKSKLEISDKIDELDAETAKLEDDKAAEENNVDEYFKEGLEKTTAA 301  
Qy 406 KKAETRLKIKITDREAESEAKKVAEEDYARSESEYNNLTQOOPKTEKPAQPSPT 464  
Db 302 KKAEE--LEKTEADLKAVNPEKPAPEBTPAPAPAPAPAPAPAPAPAPAPAPAP 355  
Qy 465 PABOPKAEKPADQAEEDYARSESEYNNLTQOOPKTEKPAQPSPT 510  
Db 356 PABOPKAEKPADQAEEDYARSESEYNNLTQOOPKTEKPAQPSPT 402



```

OY      108 NNKKREYJNKKYVESSTESOLOI---LMMESRKNQVAVSKFEDSSSSSSSSSTRPPAS 164
Db      544 EETEAEAAEEBEEAGBEEAEAEAGBEEAEAEVEEBEALIKAAAKAAKAEVEEBEAEAE 603
OY      165 DTAKPNKPTPEGKVAEAKKVY-----EAAKKAQDQEDRRNYPIT---YKTTLE 214
Db      604 E-----EBEAEEBEEVAETKEEVEAEAEVEEBEAAEEAEAEAEVEEVTSSKAKTOAB 659
OY      215 IASDVEKKKKELEJ-----YKVKANPPREQIKQAAEAVESKQAEATLKKI- 263
Db      660 VEEBEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 718
OY      264 KTDREAEAEAEARRADAKEOGPKGRAKRGVPELATPDKKENDAKSSDSVGBETLSP 323
Db      719 KSDAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 775
OY      324 SLKPEKKYAEAEAKVYEAQKKAQDQEDRRNYPINTYKTTLELAEADVEYKKKKELELV 383
Db      776 DEKPEEVEKESAEAPAPAEPAKPAPEPKPAAPKKAAPAK---VESPTSEDEPK---AAEV 828
OY      384 KEAEKEPNEBEKVQAKAEVSKKAAEATRLKIKIDRKKAEBEAKKAAEEDKVYKEKPAE 443
Db      829 EKKGAELAPKPKAPKAAAKKKEKAPVE-----KEEPEESPTBEPPKPPAA---AKKAP 878
OY      444 QP-QPAPAPKA--EKPAPA-----EKPENPAEQPAEKPRADQQAEDY-----483
Db      879 APAKPKPAPKAEAEKPEPAKPAQAKPAPAAAEEBEEDKEDDEEBEEVEAEVAPKPAEKVK 938
OY      484 ---ARSEEVNRLTQOOPKTE-KPAOP 508
Db      939 SKPAPAKEEDPEKAKOPPKPKKRPAP 967

```

RESULT 7  
 OFHUH  
 neurofilament triplet H protein - human  
 N:Alternate names: neurofilament protein, 112K  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jun-2000  
 C:Accession: S00979  
 EMBL J. 7, 1947-1955, 1988  
 A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)  
 A:Reference number: S00979, MUID:88328581, PMID:3138108  
 A:Accession: S00979  
 A:Molecule type: DNA  
 A:Residues: 1-1020 <LBE>  
 A:Cross-references: EMBL:X15306; NID:g35028; PIDD:CAA33366.1; PID:gl844130  
 A>Note: it is uncertain whether Met-1 or Met-2 is the initiator  
 C:Genetics:  
 A:Gene: GDB:NEFH  
 A:Cross-references: GDB:120225; OMIM:162230  
 A:Map position: 22q12.1-22q13.1  
 A:Introns: 295/1; 361/3; 403/2  
 C:Superfamily: neurofilament triplet H protein  
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein  
 F:1-100/Domain: amino-terminal <NTD>  
 F:101-410/Domain: rod #status predicted <ROD>  
 F:411-1020/Domain: carboxyl-terminal <CTD>  
 F:502-826/Region: 14-residue repeats  
 F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,646,652,658,666,674,682,690,696,704,712,720,728,736,744,752,760,768,776,784,792,800,808,816,824,832,840,848,856,864,872,880,888,896,904,912,920,928,936,944,952,960,968,976,984,992,1000,1008,1016,1020,1028,1036,1044,1052,1060,1068,1076,1084,1092,1100,1108,1116,1124,1132,1140,1148,1156,1164,1172,1180,1188,1196,1204,1212,1220,1228,1236,1244,1252,1260,1268,1276,1284,1292,1300,1308,1316,1324,1332,1340,1348,1356,1364,1372,1380,1388,1396,1404,1412,1420,1428,1436,1444,1452,1460,1468,1476,1484,1492,1500,1508,1516,1524,1532,1540,1548,1556,1564,1572,1580,1588,1596,1604,1612,1620,1628,1636,1644,1652,1660,1668,1676,1684,1692,1700,1708,1716,1724,1732,1740,1748,1756,1764,1772,1780,1788,1796,1804,1812,1820,1828,1836,1844,1852,1860,1868,1876,1884,1892,1900,1908,1916,1924,1932,1940,1948,1956,1964,1972,1980,1988,1996,2004,2012,2020,2028,2036,2044,2052,2060,2068,2076,2084,2092,2100,2108,2116,2124,2132,2140,2148,2156,2164,2172,2180,2188,2196,2204,2212,2220,2228,2236,2244,2252,2260,2268,2276,2284,2292,2300,2308,2316,2324,2332,2340,2348,2356,2364,2372,2380,2388,2396,2404,2412,2420,2428,2436,2444,2452,2460,2468,2476,2484,2492,2500,2508,2516,2524,2532,2540,2548,2556,2564,2572,2580,2588,2596,2604,2612,2620,2628,2636,2644,2652,2660,2668,2676,2684,2692,2700,2708,2716,2724,2732,2740,2748,2756,2764,2772,2780,2788,2796,2804,2812,2820,2828,2836,2844,2852,2860,2868,2876,2884,2892,2900,2908,2916,2924,2932,2940,2948,2956,2964,2972,2980,2988,2996,3004,3012,3020,3028,3036,3044,3052,3060,3068,3076,3084,3092,3100,3108,3116,3124,3132,3140,3148,3156,3164,3172,3180,3188,3196,3204,3212,3220,3228,3236,3244,3252,3260,3268,3276,3284,3292,3300,3308,3316,3324,3332,3340,3348,3356,3364,3372,3380,3388,3396,3404,3412,3420,3428,3436,3444,3452,3460,3468,3476,3484,3492,3500,3508,3516,3524,3532,3540,3548,3556,3564,3572,3580,3588,3596,3604,3612,3620,3628,3636,3644,3652,3660,3668,3676,3684,3692,3700,3708,3716,3724,3732,3740,3748,3756,3764,3772,3780,3788,3796,3804,3812,3820,3828,3836,3844,3852,3860,3868,3876,3884,3892,3900,3908,3916,3924,3932,3940,3948,3956,3964,3972,3980,3988,3996,4004,4012,4020,4028,4036,4044,4052,4060,4068,4076,4084,4092,4100,4108,4116,4124,4132,4140,4148,4156,4164,4172,4180,4188,4196,4204,4212,4220,4228,4236,4244,4252,4260,4268,4276,4284,4292,4300,4308,4316,4324,4332,4340,4348,4356,4364,4372,4380,4388,4396,4404,4412,4420,4428,4436,4444,4452,4460,4468,4476,4484,4492,4500,4508,4516,4524,4532,4540,4548,4556,4564,4572,4580,4588,4596,4604,4612,4620,4628,4636,4644,4652,4660,4668,4676,4684,4692,4700,4708,4716,4724,4732,4740,4748,4756,4764,4772,4780,4788,4796,4804,4812,4820,4828,4836,4844,4852,4860,4868,4876,4884,4892,4900,4908,4916,4924,4932,4940,4948,4956,4964,4972,4980,4988,4996,5004,5012,5020,5028,5036,5044,5052,5060,5068,5076,5084,5092,5100,5108,5116,5124,5132,5140,5148,5156,5164,5172,5180,5188,5196,5204,5212,5220,5228,5236,5244,5252,5260,5268,5276,5284,5292,5300,5308,5316,5324,5332,5340,5348,5356,5364,5372,5380,5388,5396,5404,5412,5420,5428,5436,5444,5452,5460,5468,5476,5484,5492,5500,5508,5516,5524,5532,5540,5548,5556,5564,5572,5580,5588,5596,5604,5612,5620,5628,5636,5644,5652,5660,5668,5676,5684,5692,5700,5708,5716,5724,5732,5740,5748,5756,5764,5772,5780,5788,5796,5804,5812,5820,5828,5836,5844,5852,5860,5868,5876,5884,5892,5900,5908,5916,5924,5932,5940,5948,5956,5964,5972,5980,5988,5996,6004,6012,6020,6028,6036,6044,6052,6060,6068,6076,6084,6092,6100,6108,6116,6124,6132,6140,6148,6156,6164,6172,6180,6188,6196,6

```

OY 60 GEOPKJDSDEKDA--REVEEYVK-KIVGVSJASTKRIHTITVALVNLNNIKNEYJAN 116
Db 497 GEETKSPPAEAAASPEKAKSPVKEEKAAPSAAKSPKEEKAAPSAAVYKSPKASP--- 553
OY 117 KIVESTSESOLOIIMMESRSKVDAAVSKFEKDSOSSSSDSSTKPEASDTAKPNKP-EP 175
Db 554 --AAKEAASPEPEAKSPKEEKAAPSAAVYKSPKAKSPAAKEAASPAEAKSPKAKSPVKBE 611
OY 176 GEKVAEAKKTV--EEAKKAKDOKEEDRNNYPITTYLTLELIAESDVEYKAELELVKVA 234
Db 612 AKSPAEEAKSPVKEEKAAPSAAVYKSPKAKSP-----KEEKAAPSKEKASPE---KEBA 660
OY 235 NEPDEQOKIKAAEAIVESK-----QAEKTRLKIKITD--REAEAEBAKRAADAKGCKPRG 288
Db 661 KSPAEEKASPVAAEAKSPKAKSPVAAEAKSPKAKSPVKEEKAAPSPEKAKSPVKEEKASPE 720
OY 289 RAKGVSPPELITPOKENDAKSDSSVGEETLPSPSLRPEKKVAEAAKKEVEAKKAKEAO 348
Db 721 KAKSPVKEEAKTPEAKKSPVKEEAKS-----PEAKASPE-KAKTLDVKSPPEAKTPEAKE 773
OY 349 KEEDRNNYPITTYLTLELIAESD-----VEYKAELEL-VKEEAKSPNNE----- 394
Db 774 ARSPADKRPEKAKSPVKEEVKSPKAKSPLKADAPAKPEKIPKGBVKSPEVKEEKPOEV 833
OY 395 KVKQAKAEVESKKAAPATRLKIKITDRKAAEBEAKKAABEDKVYK-----440
Db 834 KVKPEPPKAAEBEKAAPATPKTEEEKDSKK--EEAPKKEAPKPVKEEKEKKEPAVEKPKESKYE 891
OY 441 ----PAEPOPAAPAKAKP-----APAKPEPAAPQAPKE 472
Db 892 AKKEEAEBEKKVPPPEKEAPAKVEVEDAKPKTEKTEVAKKEPDDAKAKESPSPKAEKKEA- 950
OY 473 KPADOQAEDYVARSEEEYNRLTOOQPPKTEKPA-----QBSTP 511
Db 951 APEKKDTYEEKAKKPEEK-----PKTAAKAKEDDKTLSKESKSP 989

```

Query Match	14.2%	Score 361	DB 2	Length 1701
Best Local Similarity	29.1%	Pred. Mis. 6.2e-07		
Matches 136	Conservative	76	Mismatches 171	Indels 84
			Gaps 17	
Query	56	QAEQGEQPKLUDSPRDKAR--KEVEEYVKKIVGESYKSTKTKRHTITVALVNLNLNI---	110	
		.....		
Db	1107	EAQQAAMAKKAEEKKAEAVKKAEEBKRIIEAKKAEEBKRIIEAKKAEEBKRIIEAE	1166	
		.....		
Qy	111	-KNEVYLNKIVESTESQQLIMNESRSKVDVAVSKEPKDSSSSSSSDSTKPEASDTA-K	168	
		.....		
Db	1167	KKAEEBKRIIEAKKAEEBKRIIEAKKAEEBKRIIEAKKAEEBKRIIEAKKAEEBKRIIEAE	1223	
		.....		
Qy	169	PNKPTGEGERVAEAKKKVVEEAKKAKOKEE-----DRNRPITYTKLDELIAASDV	220	
		.....		
Db	1224	AKKAEEAKKAEEKKKAEEAKKAEEAKKAEEAKKAEEAKKAEEAKKAEEAKKAEEAKKAEEAKKAEE	1281	
		.....		







Db 737 RKAQEEBAARKQAEEBAARKQ-----AEEBAARKQAEEBAARKQAEE--- 778  
Qy 393 EEKTKQAQAEVSESKAETRLLEKI--KTDRKKAEEBAARKQAEEBDKVEKEPAEQPAPAPA 450  
Db 779 EAARKQAEEBAARKQAEEBAARKQAEEBAARKQAEEBAARKQAEEBAAR--KQAE----- 830  
Qy 451 PKAEKPPAPKPPENPAEQPAEKPADQQAEDYARRSEENRRLTQOO 498  
Db 831 -----EBAARKQAEEBAARKQAEEBAARKQAEEBAARKQAEEBAAR 866  
  
RESULT 13  
T34418  
hypothetical protein F12F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C/Accession: T34418  
R:Fulton, B.; Wohldmann, P.  
submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of C. elegans cosmid F12F3.  
A:Reference number: Z21521  
A/Accession: T34418  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3488 <FUL>  
A:Cross-references: EMBL:U80022; PIDN:ACG25885.1; GSPDB:GN00023; CESP:F12F3.3  
A:Experimental source: strain Bristol N2; clone F12F3  
C/Genetics:  
A:Gene: CESP:F12F3.3  
A:Map position: 5  
A:Intons: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 13.0%; Score 329.5; DB 2; Length 3488;  
Best Local Similarity 26.8%; Pred. No. 1.9e-05;  
Matches 158; Conservative 93; Mismatches 199; Indels 139; Gaps 28;  
  
Qy 3 ASKSEKRVHYSI-----RKFSVGASV--VASLWGVSHATENEGATQV--PTSSNRAE 54  
Db 899 AAKSKVSESVKGEKKTAKGKTYQVESEPTSKTIDTKDVGATHPADETPKKIIK 958  
Qy 55 SQAEOGR-----QPKLDSERDKARKEVEYKIVGESYA-----KSTKRHTTVALVN 105  
Db 959 KKTESDSSISQKATSEKVSQKQDEPTKPAVSEIQWTEADKSKQKET----- 1011  
Qy 106 EUNIKREYLINKIVESTESQQLIMMESRSKVD--EAVSKFEKDSSSS----- 153  
Db 1012 -----DEKTLDAEIAKTKQE--ADEKSKLDAQEKIKKVEDDARKEKELNDKTL 1062  
Qy 154 SSDSSTPEASDTPAKPKPTPEGEKVAEAKKVEEAKKQKQEDERNYPTITYKTLEL 213  
Db 1063 ESEIATTKASADKLK-----EEOQAQ--KKAEEVNAKKQKQEKDQ-----LKLDY 1107  
Qy 214 EIASDVEVKKALELVKVNAPRDEQIKQAEEVESKQAETRLKIKTDREEA--- 270  
Db 1108 EAASKKAQAELTELE-----KQAQIKKAGADAVKKQKELDENKTLKANKKSAAG 1158  
Qy 271 ---EE--ARRADAKQKPKGKRGVPEGLATPPKENDAKSSSVGEETLPSPSL 325  
Db 1159 LKTEESAAKSKQVEQAKLDAQTK--AKTAEKQT--KLEDEKSTTESSEKTEVDE--- 1212  
Qy 326 KPEKRV-----AAEKKVVEA--KKKADQKEED--RRNPPTNYKTLEL 366  
Db 1213 KPKKTKLKTTEKSDSSISQKSETSKTVESAGSESETQVADAAKQKETDEKQKLEA 1272  
Qy 367 EIASDVEVKKALELVKEEAPRNEKVKQAQAEVESKKAETRLKIKTDKKAEE 426  
Db 1273 EIT-----AKKSADKSKLEA-----ESKLKKA--AEVBAKKQKQEKDQKLDTEAASK 1321  
Qy 427 AK-----RKAABEDKV--KEKPAEQPPAPAPKAEKPPAPKPPENPAEQPA-- 471  
Db 1322 AAEEKLEKQSHIKKKAEEVDAVKKQKLEKQRLSEBAATKKADAEKTLKEQKKKAEE 1381  
Qy 472 -----EKPADQQA--EDVARRSEEEYNRLTQOQPPKTEKPAQPS 509

Db 1382 IALIEIQKEQEKLAQEQSRLEDEAKKSAEKQKLESETSKQTEBAPKES 1430  
  
RESULT 14  
S51364  
sperm tail-specific protein msl101(2) - fruit fly (Drosophila hydei)  
C:Species: Drosophila hydei  
C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 17-Mar-1999  
C/Accession: S51364; S34154  
R:Neesen, U.; Padmanabhan, S.; Buenemann, H.  
Eur. J. Biochem. 225, 1089-1095, 1994  
A>Title: Randomly arranged repeats of a novel highly charged 16-amino-acid motif repre-  
sents a helical rod within the extremely elongated spermatozoa of Drosophila hydei.  
A/Reference number: S51364; MUID:55045538; PMID:7557199  
A/Accession: S51364  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1390 <NE>  
A:Cross-references: EMBL:X73481  
R:Neesen, U.; Heinlein, U.A.O.; Buenemann, H.  
submitted to the EMBL Data Library, June 1993  
A/Reference number: S34153  
A/Accession: S34154  
A:Molecule type: DNA  
A:Residues: 1-163; 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>  
A:Cross-references: EMBL:X73481; NID:9313201; PID:9313202  
C/Genetics:  
A:Gene: msl101(2)  
A:Cross-references: FlyBase:FBgn0011816

Query Match 12.9%; Score 328; DB 2; Length 1390;  
Best Local Similarity 27.0%; Pred. No. 9.4e-06;  
Matches 140; Conservative 79; Mismatches 220; Indels 80; Gaps 15;  
  
Qy 38 ENEGATQVPTSSNRAE--SQAEOGRPKLDSER--DKARKEVEYKIVGESYAKST 93  
Db 362 EKAQCELAQKKEADKCKCEBAANKKQAEEKQKQKAEKKAEEKKEAAKKE 421  
Qy 94 KKRHTTVALVNLINIKNEYLINKIVESTESQQLIMMESRSKVDVAVSFEKDSSSS 153  
Db 422 KE-----AAERKCEBELANIKKAIEKK--KCEBAKKEKAAERKCEBELAKK 468  
Qy 154 SSDSSTPEASDTPAKPKPTPEGEKVAEAKKVEEAKKQKQEDERNYPTITYKTLE 212  
Db 469 IKKAERKKECEETAKKQKQVEAERKCEBELAKKIKKQKQKQKQKQKQKQKQK 528  
Qy 213 -----LEIAESD-----VEYKAELELVKVNAPRDEQIKQA-----EAEV 250  
Db 529 AAKKREAAKCKKCEAAKKAERKKAERKKEKSAKRR--EAAEKKCEAAKKEAAE 586  
Qy 251 ESKQAETRLKIKTDREEAEEBAKRRADAKQKQKPKRAK-----RGVPELATPDK 303  
Db 587 KKKCEAAKKEKEVAERKCEBELAKKIKKAEEKKCEAAKKEKAAERKCEBELAKK 646  
Qy 304 KENDAKSSSVSEBELPSPSLPE-----KKVAEAKVVEA--KKKADQK 349  
Db 647 KAAEKKKCKKLAKKEKETAKKCEAAKKAERKKAERKKAERKKAERKKAERKKA 706  
Qy 350 EEDRRNYPNTYKTLELEIAESDVEVKKALELVKEEAPRNEKVKQAQAEVESKAE 409  
Db 707 KKEKAAERKKEBELAKKIK-----KAAEKKCKKLAKKQKQKQKQKQKQKQKQ 759  
Qy 410 ATLEKIKTDKKAEEBAKKAEEBKVEKEPAEQPAPAPKAEKPPAPKPPENPAEQ 469  
Db 760 -----LKEKKCRELAKKKAEEKCKKCEAAKKEKKAEEKCKKCEYAKKKEE--AEKK 810  
Qy 470 KAEKPADQQAEDYARRSE--EYNRLTQOQPPKTEKPAQ 507  
Db 811 KCEKTAKKREAAEKKCEAAKKAERKKEAEKKEKCKCEYAK 849  
  
RESULT 15

protein UNC-89 - *Caenorhabditis elegans*  
 r29757  
 C.Species: *Caenorhabditis elegans*  
 C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999  
 C.Accession: r29757  
 R.Du, Z.; Le, T.T.; Wilson, R.  
 submitted to the EMBL Data Library, May 1997  
 A.Description: The sequence of C. elegans cosmid C09D1.  
 A.Reference number: Z20679  
 A.Accession: r29757  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-6642 <DUZ>  
 A.Cross-references: EMBL:AF003111; PIDD:AA854132.1; GSPDB:GN00019; CESP:unc-89  
 A.Experimental source: strain Bristol N2; clone C09D1  
 C.Genetics:  
 A.Gene: CESP:unc-89  
 A.Map position: 1  
 A.Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 602/1; 6027/1; 6061/3; 6155/2; 6515/1; 6552/3; 6609/1  
 3;

Query Match	12.9%;	Score 328;	DB 2;	Length 6642;
Best Local Similarity	26.9%;	Pred. No. 3.9e-05;		
Matches 133;	Conservative 76;	Mismatches 203;	Indels 82;	Gaps 20;

```

QY 37 IENEATQVPTSNNANESQAEO-GEQ-----PKLUDSRDQARKEVEYVAKIYESGA 90
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1411 TKKESPPSSPTKKKESPPSSPTKKTGDVEKESPPYSPYTKKESPEKPEDV-----KSPV 148
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 91 KSTKRRHTITVALVNLNNIKNEYLNKIIVESTSESOLQIIMRESKRYDAVSKPEKSS 150
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1485 KKEKPPDANIVEVSESETTIKTEETITMTTEMHSE-----ESRITSVAKKPTPEKVDEK 153
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 SSSSSSSSSTKPEASDTAK---DNKPTPEGEKVAE-----AKKVVEAKKADQKEEDRR 201
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1539 PKSPPTKKDKSPKESITTEIKSPVKKEKSEPEKVEEKPASPTKKESPEKSPATTKKSENEV 1596
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 202 NYPTITTYKTLEIABSDVEYKKALELVKVAANPRDQOKIQAQBAEVESQAQATRLK 261
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1599 KSPTKKESPEKSVVE-----EIKSPKESPEKAD-KEKSPTKKESPEKATEDV 1644
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 KIKTGREAE--BAKRBRADAKQOKP-----KQARQGVGELATDQKEN 306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1650 KSPTKKESPEVVEEPTSPYTKKSSPTKKTDEYVSPYTKKESPOYVEEKASPTKE- 1700
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 DAKSSDSVGEETLPSPLPEKKYAAEAKVBEAKKAQDQKEEDRRNYPNTYKTLEL 366
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1709 --KSPSEKSVBEVSPKPEKSPK--AEERPKSPYTKKESPEKSAABEVVSPYTKKESPEK 1766
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 367 EIASDVEYKKALELVK--BEAPEPRNEKVQAKAVESKKAQATRLKIKITDRKKA 423
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1765 SAEEKPKSPYTKKSSPYVMADDEVKSPYTKK--SPEVVEEKPPASPTKE--KTPPEKSA 1819
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 424 EBEAK-----RKAEEEDKV--EKPAPOQAPAPKPAEKAPAPKPPENPAQ 468
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1820 ABEELSPYTKKESPPSPYTKTGDSEKESPEKPEEKPK-SPYPA-KSPGSPK-KKSSKS 1876
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 469 PKAEKPADQAAED 482
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1877 PEAEKPPAPKLT RD 1890
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: November 21, 2003, 13:34:25  
Job time : 22.7062 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:22:21 ; Search time 19.5369 Seconds

(without alignments)  
1462.401 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442

Perfect score: 889

Sequence: 1 IKTDREAEAEAKRRADAKE.....AEEAKRKAEDKVEKPA 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_19Jun03.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	406	20	AAV49225
2	889	100.0	406	20	AAV32098
3	889	100.0	428	20	AAV49238
4	889	100.0	428	20	AAV32110
5	889	100.0	446	20	AAV49143
6	889	100.0	453	19	AAW55081
7	889	100.0	453	23	ABP54575
8	889	100.0	460	20	AAV49250
9	889	100.0	655	20	AAV49226

10	889	100.0	655	20	AAV32099	Choline binding pr
11	889	100.0	693	23	AAU76890	Pneumococcal surfa
12	889	100.0	693	24	ABU02720	S. pneumoniae type
13	889	100.0	694	21	AAV81653	Streptococcus pneu
14	886	99.7	460	20	AAV32189	N-terminal choline
15	884	99.4	446	20	AAV49140	Amino acid sequenc
16	884	99.4	459	20	AAV49251	N-terminal region
17	884	99.4	459	20	AAV32190	N-terminal choline
18	803.5	90.4	376	20	AAV49231	N-terminal truncat
19	803.5	90.4	376	20	AAV32103	Choline binding pr
20	803.5	90.4	413	20	AAV49247	N-terminal region
21	803.5	90.4	413	20	AAV32186	N-terminal choline
22	803.5	90.4	414	20	AAV49150	Amino acid sequenc
23	803.5	90.4	564	19	AAW62654	C3 binding protein
24	803.5	90.4	581	20	AAV43394	S. pneumoniae SpSa
25	803.5	90.4	631	18	AAW6444	Cbpa, a choline bi
26	803.5	90.4	663	20	AAV49232	Cbpa of serotype R
27	803.5	90.4	663	20	AAV32104	Choline binding pr
28	803.5	90.4	663	20	AAV43395	S. pneumoniae chol
29	803.5	90.4	663	23	AAE29201	Streptococcus pneu
30	803.5	90.4	663	23	ABG60639	Partial sequence o
31	803.5	90.4	663	24	ABP55319	Streptococcus pneu
32	799.5	89.9	701	23	AAU76888	Pneumococcal surfa
33	795	89.4	439	20	AAV49242	N-terminal region
34	795	89.4	439	20	AAV32181	N-terminal choline
35	793	89.2	419	20	AAV49243	N-terminal region
36	793	89.2	419	20	AAV32182	N-terminal choline
37	792	89.1	406	20	AAV49152	Amino acid sequenc
38	792	89.1	425	20	AAV49145	Amino acid sequenc
39	791.5	89.0	414	20	AAV49144	Amino acid sequenc
40	784	88.2	438	20	AAV49141	Amino acid sequenc
41	784	88.2	439	20	AAV49249	N-terminal region
42	784	88.2	439	20	AAV32188	N-terminal choline
43	775.5	87.2	412	20	AAV49151	Amino acid sequenc
44	775.5	87.2	425	20	AAV49248	N-terminal region
45	775.5	87.2	425	20	AAV32187	N-terminal choline

## ALIGNMENTS

RESULT 1	AAV49225	standard; protein; 406 AA.
ID	AAV49225	
XX	AAV49225	
AC	AAV49225	
XX	07-FEB-2000 (first entry)	
DT	07-FEB-2000	
XX		
DE	Polypeptide R2 of the N-terminal truncate of Cbpa serotype type 4.	
XX		
KM	Choline binding protein A; Cbpa; truncate; immune response; infection;	
XX	pneumococcal bacterium; vaccine.	
OS	Streptococcus pneumoniae.	
XX		
XX	WO951187-A2.	
PN	14-OCT-1999.	
PD		
XX	07-APR-1999;	99WO-US07668.
PF	07-APR-1998;	98US-0056019.
XX	07-APR-1998;	98US-0080878.
PR		
XX	(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
PA		
XX	Tuomanen EI, Measure HR;	
PI		
XX	WPI, 1999-620161/53.	
XX	N-PSDB; AA231401.	
PT	Novel polypeptides, used to develop products for the diagnosis,	

PT prevention and treatment of pneumococcal infections -  
XX  
XX Claim 2; Page 7-8; 85pp; English.  
CC The invention provides novel isolated polypeptides comprising the amino  
XX acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC AA49225, AA49227, AA49230, AA49231, AA49233 and AA49238. The  
XX polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use.  
XX  
XX Sequence 406 AA;  
SQ  
Query Match 100.0%; Score 889; DB 20; Length 406;  
Best Local Similarity 100.0%; Pred. No. 1.2e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IKTDREAEAEAEKRRADAKKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAEKRRADAKKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286  
QY 61 PSLEPEKVAEAEKVEAEAKKKAEDQKEEDRRNPTNTYKTLLEIASDVEVKAAEEL 120  
DB 287 PSLEPEKVAEAEKVEAEAKKKAEDQKEEDRRNPTNTYKTLLEIASDVEVKAAEEL 346  
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEAKRAAEEDVKYKCPA 180  
DB 347 VKEAEKPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEAKRAAEEDVKYKCPA 406  
RESULT 2  
AA432098  
ID AA432098 standard; Protein; 406 AA.  
AC AA432098;  
DT 01-FEB-2000 (first entry)  
DE Choline binding protein A (Cbpa) truncate R2.  
XX  
XX Choline binding protein; Cbpa; truncate; adhesin; immunogen;  
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
KW otitis media; pneumonia.  
XX  
XX Streptococcus pneumoniae.  
OS Synthetic.  
XX  
XX WO9951187-A2.  
XX  
XX 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US07669.  
XX  
XX 07-APR-1998; 98US-0056019.  
PR 07-APR-1998; 98US-0080878.  
XX  
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (MEDI-) MEDIMUNE INC.  
XX  
XX Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;  
XX  
XX WPI; 1999-633690/54.

DR N-PSDB; AA234451.  
XX  
XX New N-terminal choline binding protein A truncate polypeptides used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections -  
XX  
XX Claim 2; Page 85-88; 160pp; English.  
PS  
CC The present sequence represents polypeptide R2, a new N-terminal  
CC choline binding protein A (Cbpa) truncate of Streptococcus pneumoniae  
CC serotype type 4. Host-vector systems for production of R2 and other  
CC N-terminal Cbpa truncates are provided. Important features of R2  
CC include: (1) complete correlation of bioactivity with full-length  
CC Cbpa for recognition of purified glycoconjugate receptor analogues,  
CC lung cells and animal models; and (2) cross-protection between type  
CC 4 derived agents and bacteria in in vitro assays using other  
CC serotypes (e.g. 6B and 2), important for use in vaccine, prophylaxis  
CC and therapy. The invention relates generally to novel N-terminal  
CC Cbpa truncates and polynucleotides encoding them, host-vector  
CC systems, and antibodies that specifically bind to the truncates.  
CC The invention also relates to vaccines including N-terminal Cbpa  
CC truncates, which provide protection or elicit protective antibodies  
CC to bacterial infection, specifically pneumococcus, and to use of  
CC antibodies and antagonists against such polypeptides in diagnosis  
CC and passive immunotherapy. The polypeptides and/or polynucleotides  
CC are also useful as competitive inhibitors of bacterial adhesion of  
CC pneumococcus.  
XX  
XX Sequence 406 AA;  
SQ  
Query Match 100.0%; Score 889; DB 20; Length 406;  
Best Local Similarity 100.0%; Pred. No. 1.2e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IKTDREAEAEAEKRRADAKKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTDREAEAEAEKRRADAKKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286  
QY 61 PSLEPEKVAEAEKVEAEAKKKAEDQKEEDRRNPTNTYKTLLEIASDVEVKAAEEL 120  
DB 287 PSLEPEKVAEAEKVEAEAKKKAEDQKEEDRRNPTNTYKTLLEIASDVEVKAAEEL 346  
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEAKRAAEEDVKYKCPA 180  
DB 347 VKEAEKPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEAKRAAEEDVKYKCPA 406  
RESULT 3  
AA49238  
ID AA49238 standard; Protein; 428 AA.  
AC AA49238;  
DT 07-FEB-2000 (first entry)  
DE Choline binding protein A (Cbpa) serotype type 4 amino acid sequence.  
XX  
XX Choline binding protein A; Cbpa; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
XX  
XX Streptococcus pneumoniae.  
OS  
XX  
XX WO9951187-A2.  
XX  
XX 14-OCT-1999.  
XX  
XX 07-APR-1999; 99WO-US07668.  
XX  
XX 07-APR-1998; 98US-0056019.  
PR 07-APR-1998; 98US-0080878.  
XX  
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA

PI Tuomanen EI, Maure HR;  
 XX WPI; 1999-620161/53.  
 DR Novel polypeptides, used to develop products for the diagnosis,  
 XX prevention and treatment of pneumococcal infections  
 XX Claim 7, Page 7, 85pp; English.  
 XX  
 CC The invention provides novel isolated polypeptides comprising the amino  
 CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
 CC (CbpaT). The polypeptides can be selected from sequences shown in  
 CC AAY9225, AAY9227, AAY9230, AAY9231, AAY9233 and AAY9238. The  
 CC polypeptides can be used for inducing an immune response in a subject  
 CC which has been exposed to or infected with a pneumococcal bacterium. They  
 CC can also be used for preventing infection by a pneumococcal bacterium.  
 CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
 CC used for treating a subject infected with or exposed to a pneumococcal  
 CC bacterium. Antibodies specifically binding the polypeptides can be used  
 CC for detection and diagnosis and for preventing pneumococcal attachment to  
 CC a mucosal surface. The products can be used in humans and other animals  
 CC such as domestic animals, such as feline or canine subjects, farm animals  
 CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
 CC animals (whether in the wild or in a zoological garden), research  
 CC animals, such as mice, rats, rabbits, goats, sheep, pigs, cats,  
 CC i.e. for veterinary medical use.  
 CC  
 SQ Sequence 428 AA;  
 Query Match 100.0%; Score 889; DB 20; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-62;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 IKTRBEAEBAEKRRADAKGKPKGKRGVPGELATPPDKENDAKSSDSSVGEETLPS 60  
 Db 227 IKTRBEAEBAEKRRADAKGKPKGKRGVPGELATPPDKENDAKSSDSSVGEETLPS 286  
 Oy 61 PSLKPEKKVAEAEKKVEAKKKAEQKEDRRNPTNTYKTLLELEIASDVEVKKAELEL 120  
 Db 287 PSLKPEKKVAEAEKKVEAKKKAEQKEDRRNPTNTYKTLLELEIASDVEVKKAELEL 346  
 Oy 121 VKEBAKPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
 Db 347 VKEBAKPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 4  
 AAY32110  
 ID AAY32110 standard; Protein; 428 AA.  
 XX  
 AC AAY32110;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Choline binding protein A (Cbpa) N-terminal truncate.  
 XX  
 KW Choline binding protein; Cbpa; truncate; adhesin; immunogen;  
 KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
 KW otitis media; pneumonia.  
 XX  
 OS Streptococcus pneumoniae.  
 OS Synthetic.  
 XX  
 PN W09951188-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-US07669.  
 XX  
 PR 07-APR-1998; 98US-0056019.  
 PR 07-APR-1998; 98US-0080878.  
 XX  
 PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PA (MEDI-) MEDIMUNE INC.  
 XX  
 XX Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;  
 XX WPI; 1999-633690/54.  
 DR Novel polypeptides, used to develop products for the diagnosis,  
 XX prevention and treatment of  
 XX pneumococcal infections  
 XX Claim 7, Page 119-122; 160pp; English.  
 PS  
 CC The present sequence represents a new N-terminal choline binding  
 CC protein A (Cbpa) truncate that maintains the tertiary structure of  
 CC Streptococcus pneumoniae serotype type 4 Cbpa native protein.  
 CC Host-vector systems for production of N-terminal Cbpa truncates are  
 CC provided. The invention relates generally to novel N-terminal  
 CC Cbpa truncates and polynucleotides encoding them, host-vector  
 CC systems, and antibodies that specifically bind to the truncates.  
 CC The invention also relates to vaccines including N-terminal Cbpa  
 CC truncates, which provide protection or elicit protective antibodies  
 CC to bacterial infection, specifically pneumococcus, and to use of  
 CC antibodies and antagonists against such polypeptides in diagnosis  
 CC and passive immunotherapy. The polypeptides and/or polynucleotides  
 CC are also useful as competitive inhibitors of bacterial adhesion of  
 CC pneumococcus.  
 CC  
 SQ Sequence 428 AA;  
 Query Match 100.0%; Score 889; DB 20; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-62;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 IKTRBEAEBAEKRRADAKGKPKGKRGVPGELATPPDKENDAKSSDSSVGEETLPS 60  
 Db 227 IKTRBEAEBAEKRRADAKGKPKGKRGVPGELATPPDKENDAKSSDSSVGEETLPS 286  
 Oy 61 PSLKPEKKVAEAEKKVEAKKKAEQKEDRRNPTNTYKTLLELEIASDVEVKKAELEL 120  
 Db 287 PSLKPEKKVAEAEKKVEAKKKAEQKEDRRNPTNTYKTLLELEIASDVEVKKAELEL 346  
 Oy 121 VKEBAKPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
 Db 347 VKEBAKPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 5  
 AAY49143  
 ID AAY49143 standard; Protein; 446 AA.  
 XX  
 AC AAY49143;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Amino acid sequence of choline-binding protein fragment #7.  
 XX  
 KW Truncated surface binding protein; alpha helix; choline binding protein;  
 KW vaccine; invasive bacterial infection; otitis media; sepsis;  
 KW meningitis; lobar pneumonia infection; antibody; immature immune system;  
 KW immunocompromised.  
 XX  
 OS Streptococcus pneumoniae.  
 OS  
 XX  
 PN W09951266-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 06-APR-1999; 99WO-US07680.  
 XX  
 PR 07-APR-1998; 98US-0080878.  
 PR 15-MAY-1998; 98US-0085743.  
 XX  
 PA (MEDI-) MEDIMUNE INC.

```
XX XX
PI Wizemann TM, Koenig S, Johnson LS;
XX XX
DR WPI; 1999-601465/51.
DR N-PSDB; AA231083.
XX XX
PT New pneumococcal proteins useful as vaccines and for diagnosis of
PT pneumococcal infections -
XX XX
PS Claim 10; Page 72-73; 98pp; English.
XX XX
CC AA49137-Y49152 are amino acid sequences that are fragments of choline
CC binding proteins (CBP). The fragments of the protein are the alpha helix
CC forming parts of the CBPs from Streptococcus pneumoniae. The
CC polypeptides do not contain the actual choline binding fragment. The
CC polypeptides and the nucleotide sequences that encode them
CC (AA31077-231092) are used in the invention, which relates to polypeptide
CC truncates of a pneumococcal surface binding protein containing the highly
CC conserved immunogenic alpha helical portion and no choline binding
CC portion. The polypeptides are used as immunogens in a bacterial vaccine.
CC The vaccine can be used for preventing (immunizing) or treating invasive
CC bacterial (especially pneumococcal) infections, especially otitis media
CC (caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia
CC infections. Antibodies raised against the polypeptide are useful for
CC detection, prevention (passive immunity) and treatment of S. pneumoniae
CC infections. The vaccines are especially useful in immunocompromised
CC patients, those with an immature immune system, or patients with an on
CC going pneumococcal infection. The vaccine avoids unnecessary expense and
CC provides broad protection against a range of pneumococcal serotypes and
CC it produces an improved and enhanced effect in preventing bacterial
CC infections.
XX XX
SQ Sequence 446 AA;
XX XX
Query Match 100.0%; Score 889; DB 20; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IKTDREBEAEEBAKRRADAKKEQKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 228 IKTDREBEAEEBAKRRADAKKEQKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 287
OY 61 PSLKPEKKVAEAEKKVVEAEKKKAEQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 120
DB 268 PSLKPEKKVAEAEKKVVEAEKKKAEQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 347
OY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRRKAAEDVKKEKPA 180
DB 348 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRRKAAEDVKKEKPA 407
XX XX
RESULT 6
AAM55081
ID AAM55081 standard; Protein; 453 AA.
XX AC AAM55081;
XX XX
DT 02-OCT-1998 (first entry)
XX XX
DE Streptococcus pneumoniae SP0023 protein.
XX XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX XX
OS Streptococcus pneumoniae.
XX OS
XX PN MO9818930-A2.
XX XX
XX PD 07-MAY-1998.
XX XX
XX PF 30-OCT-1997; 97WO-US19422.
XX XX
XX PR 31-OCT-1996; 96US-0029960.
```

```
XX XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27342.
XX XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
XX XX
PS Claim 11; Page 56; 118pp; English.
XX XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX XX
SQ Sequence 453 AA;
XX XX
Query Match 100.0%; Score 889; DB 19; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.4e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IKTDREBEAEEBAKRRADAKKEQKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREBEAEEBAKRRADAKKEQKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 84
OY 61 PSLKPEKKVAEAEKKVVEAEKKKAEQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 120
DB 85 PSLKPEKKVAEAEKKVVEAEKKKAEQKEDRRNYPNTYTKTLELEIASDVEVKKAEL 144
OY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRRKAAEDVKKEKPA 180
DB 145 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRRKAAEDVKKEKPA 204
XX XX
RESULT 7
ABP54575
ID ABP54575 standard; Protein; 453 AA.
XX AC ABP54575;
XX XX
DT 04-SEP-2002 (first entry)
XX XX
DE S. pneumoniae SP023 protein sequence SEQ ID NO:38.
XX XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX XX
OS Streptococcus pneumoniae.
XX OS
XX PN US2002061545-A1.
XX XX
XX PD 23-MAY-2002.
XX XX
XX PF 22-JAN-2001; 2001US-0765272.
XX XX
XX PR 30-OCT-1997; 97US-0961083.
XX XX
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
```

PA (BARA/) BARASH S C.  
PA (DILL/) DILLON P J.  
PA (DOUG/) DOUGHERTY B.  
PA (FANN/) FANNON M R.  
PA (ROSE/) ROSEN C A.  
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;  
XX WPI: 2002-479261/51.  
XX N-PSDB: ABQ84810.  
XX  
XX New Streptococcus pneumoniae antigens, useful for detecting  
XX Streptococcus and for preventing or attenuating disease caused by  
XX Streptococcus infection -  
XX  
XX Claim 11, Page 24-25; 70pp; English.  
XX  
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
XX The S. pneumoniae antigens have antibacterial activity and can be  
XX used in vaccines. The S. pneumoniae antigens can also be used to  
XX prevent or attenuate a Streptococcal infection in an animal. The  
XX polynucleotides encoding the S. pneumoniae antigens can be used to  
XX detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
XX which are used in an example from the present invention.  
SQ Sequence 453 AA;

Query Match 100.0%; Score 889; DB 23; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1,4e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERARRADAKGKPKGRKRGVPGELATPDCKENDAKSSDSSVGEETLPS 60  
DB 25 IKTDREAEAEERARRADAKGKPKGRKRGVPGELATPDCKENDAKSSDSSVGEETLPS 84  
QY 61 PSUKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKALEL 120  
DB 85 PSUKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKALEL 144  
QY 121 VKEAKEPRNEEKVKAKAEVESKKAETRLKIKTDRKKAEEBAKKAABEDKVKKEKPA 180  
DB 145 VKEAKEPRNEEKVKAKAEVESKKAETRLKIKTDRKKAEEBAKKAABEDKVKKEKPA 204

RESULT 8  
AAV49250  
ID AAV49250 standard; Protein; 460 AA.

XX  
AC AAV49250;

DT 07-FEB-2000 (first entry)

XX N-terminal region of CbpA polypeptide Ntype4.

XX Choline binding protein A; CbpA; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
XX  
OS Streptococcus pneumoniae.

XX  
PN WO9951187-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99WO-US07668.

PR 07-APR-1998; 98US-0056019.

XX 07-APR-1998; 98US-0080878.

XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX

PI Tuomanen EI, Masure HR;  
XX  
XX WPI: 1999-620161/53.  
XX  
XX Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections -  
XX  
XX Disclosure; Fig 2A-B; 85pp; English.  
XX  
XX The invention provides novel isolated polypeptides comprising the amino  
XX acid sequence of an N-terminal choline binding protein A (CbpA) truncate  
XX (CbpAT). The polypeptides can be selected from sequences shown in  
XX AAV49225, AAV49227, AAV49230, AAV49231, AAV49233 and AAV49238. The  
XX polypeptides can be used for inducing an immune response in a subject  
XX which has been exposed to or infected with a pneumococcal bacterium. They  
XX can also be used for preventing infection by a pneumococcal bacterium.  
XX Vaccines comprising the polypeptides or encoding nucleic acids can be  
XX used for treating a subject infected with or exposed to a pneumococcal  
XX bacterium. Antibodies specifically binding the polypeptides can be used  
XX for detection and diagnosis and for preventing pneumococcal attachment to  
XX a mucosal surface. The products can be used in humans and other animals  
XX such as domestic animals, such as feline or canine subjects, farm animals  
XX such as bovine, equine, caprine, ovine, and porcine subjects, wild  
XX animals (whether in the wild or in a zoological garden), research  
XX animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
XX i.e. for veterinary medical use.  
SQ Sequence 460 AA;

Query Match 100.0%; Score 889; DB 20; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1,4e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERARRADAKGKPKGRKRGVPGELATPDCKENDAKSSDSSVGEETLPS 60  
DB 242 IKTDREAEAEERARRADAKGKPKGRKRGVPGELATPDCKENDAKSSDSSVGEETLPS 301  
QY 61 PSUKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKALEL 120  
DB 302 PSUKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKALEL 361  
QY 121 VKEAKEPRNEEKVKAKAEVESKKAETRLKIKTDRKKAEEBAKKAABEDKVKKEKPA 180  
DB 362 VKEAKEPRNEEKVKAKAEVESKKAETRLKIKTDRKKAEEBAKKAABEDKVKKEKPA 421

RESULT 9  
AAV49226  
ID AAV49226 standard; Protein; 655 AA.

XX  
AC AAV49226;

DT 07-FEB-2000 (first entry)

XX CbpA of serotype 4 amino acid sequence.

XX Choline binding protein A; CbpA; truncate; immune response; infection;  
KW pneumococcal bacterium; vaccine.  
XX  
OS Streptococcus pneumoniae.

XX  
PN WO9951187-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99WO-US07668.

PR 07-APR-1998; 98US-0056019.

XX 07-APR-1998; 98US-0080878.

XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Tuomanen EI, Masure HR;

XX WPI: 1999-620161/53.  
DR N-PSDB; AA231402.  
XX  
PT Novel polypeptides, used to develop products for the diagnosis,  
PT prevention and treatment of pneumococcal infections  
XX  
PS Disclosure; Page 9; 85pp; English.  
XX  
CC The invention provides novel isolated polypeptides comprising the amino  
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate  
CC (CbpaT). The polypeptides can be selected from sequences shown in  
CC (AA9225, AA9227, AA9230, AA9231, AA9232 and AA9238). The  
CC polypeptides can be used for inducing an immune response in a subject  
CC which has been exposed to or infected with a pneumococcal bacterium. They  
CC can also be used for preventing infection by a pneumococcal bacterium.  
CC Vaccines comprising the polypeptides or encoding nucleic acids can be  
CC used for treating a subject infected with or exposed to a pneumococcal  
CC bacterium. Antibodies specifically binding the polypeptides can be used  
CC for detection and diagnosis and for preventing pneumococcal attachment to  
CC a mucosal surface. The products can be used in humans and other animals  
CC such as domestic animals, such as feline or canine subjects, farm animals  
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild  
CC animals (whether in the wild or in a zoological garden), research  
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,  
CC i.e. for veterinary medical use.  
XX  
SQ Sequence 655 AA;  
Query Match 100.0%; Score 889; DB 20; Length 655;  
Best Local Similarity 100.0%; Pred. No. 2.1e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 IKTRBEAEAEKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTRBEAEAEKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAEKKAEDOKEDRRNYPNTYKTLLEIAESDVEVKAELEL 120  
DB 287 PSLEPEKVAEAEKKEVEAEKKAEDOKEDRRNYPNTYKTLLEIAESDVEVKAELEL 346  
OY 121 VKEAEKEPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAEKEPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406  
RESULT 10  
AA923099  
ID AA923099 standard; Protein: 655 AA.  
XX  
AC AA923099;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Choline binding protein A (Cbpa).  
XX  
KW Choline binding protein; Cbpa; adhesin; immunogen;  
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;  
KW otitis media; pneumonia.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9951188-A2.  
XX  
PD 14-OCT-1999.  
XX  
PF 07-APR-1999; 99WO-US07669.  
XX  
PR 07-APR-1998; 98US-0056019.  
XX  
PR 07-APR-1998; 98US-0080878.  
XX  
PA (SJDUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX  
PA (MEDI-) MEDIMUNE INC.

XX  
PI Tuomannen EI, Maure HR, Witzemann TM, Johnson LS, Koenig S;  
XX  
DR WPI: 1999-633690/54.  
XX  
DR N-PSDB; AA234452.  
XX  
PT New N-terminal choline binding protein A truncate polypeptides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT pneumococcal infections  
XX  
PS Disclosure; Page 88-93; 160pp; English.  
XX  
CC This sequence represents the choline binding protein A (Cbpa)  
CC of Streptococcus pneumoniae serotype type 4. The invention provides  
CC novel N-terminal Cbpa truncated polypeptides (see AA923098-110 and  
CC AA923179-90) and polynucleotides encoding them, host-vector systems,  
CC and antibodies that specifically bind to the polypeptides. The  
CC invention also relates to vaccines including the polypeptides,  
CC which provide protection or elicit protective antibodies to  
CC bacterial infection, specifically pneumococcus, and to antibodies  
CC and antagonists against such polypeptides for use in diagnosis and  
CC passive immunotherapy. The polypeptides and/or polynucleotides are  
CC also useful as competitive inhibitors of bacterial adhesion of  
CC pneumococcus.  
XX  
SQ Sequence 655 AA;  
Query Match 100.0%; Score 889; DB 20; Length 655;  
Best Local Similarity 100.0%; Pred. No. 2.1e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 IKTRBEAEAEKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
DB 227 IKTRBEAEAEKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 286  
OY 61 PSLEPEKVAEAEKKEVEAEKKAEDOKEDRRNYPNTYKTLLEIAESDVEVKAELEL 120  
DB 287 PSLEPEKVAEAEKKEVEAEKKAEDOKEDRRNYPNTYKTLLEIAESDVEVKAELEL 346  
OY 121 VKEAEKEPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180  
DB 347 VKEAEKEPRNEEKYKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406  
RESULT 11  
AAU76890  
ID AAU76890 standard; Protein: 693 AA.  
XX  
AC AAU76890;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pneumococcal surface protein C19 TIGR (PspC19 TIGR).  
XX  
KW H1c; factor H; FH; antibacterial; PspC19 TIGR;  
KW pneumococcal surface protein C19 TIGR.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200208426-A2.  
XX  
PD 31-JAN-2002.  
XX  
PF 19-JUL-2001; 2001WO-EP08409.  
XX  
PR 20-JUL-2000; 2000SE-0002738.  
XX  
PA (HANS-) HANSA MEDICAL AB.  
XX  
PI Björck L, Sjöholm A, Janulczyk R, Pozzi G, Iannelli F;  
XX  
DR WPI: 2002-257337/30.



PT Polypeptide having ability to bind factor H, and proteins and peptides  
PT derived from polypeptides useful as vaccine for infections caused by  
PT bacteria and for identifying agents inhibiting binding of factor H to  
PT protein

PS Disclosure: Page 59-61, 63pp; English.

XX The invention relates to the Streptococcus pneumoniae Hic polypeptide  
CC having the ability to bind factor H (fH). The Hic protein and the  
CC polynucleotide encoding it can be used to make vaccine compositions  
CC capable of generating an immune response to Streptococcus pneumoniae or  
CC binding to an anti-protein Hic antibody. The protein and its homologues  
CC are useful for identifying an agent which inhibits binding of factor H to  
CC Streptococcus pneumoniae comprising incubating any one of the  
CC polypeptides with factor H and a test agent, monitoring binding of fH to  
CC the polypeptide and determining thereby whether the test agent inhibits  
CC binding of factor to the polypeptide. This sequence represents the  
CC pneumococcal surface protein C19 TIGR (PepC19 TIGR), a factor H binding  
CC protein.

XX Sequence 693 AA;

Query Match 100.0%; Score 889; DB 23; Length 693;  
Best Local Similarity 100.0%; Pred. No. 2.3e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEARRADAKGKRGVPEGLATPPDKENDASSDSVGEETLPS 60  
DB 265 IKTDREAEAEARRADAKGKRGVPEGLATPPDKENDASSDSVGEETLPS 324  
QY 61 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAEL 120  
DB 325 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAEL 384  
QY 121 VKEBAKEPRNEEKVQAKAEVESKKAATRLLEKIKTDKKAAEEBAKKAEDVKKEKPA 180  
DB 385 VKEBAKEPRNEEKVQAKAEVESKKAATRLLEKIKTDKKAAEEBAKKAEDVKKEKPA 444

RESULT 12

ABU02720  
ID ABU02720 standard; Protein; 693 AA.

XX AC ABU02720;

XX DT 11-FEB-2003 (first entry)

XX DE 5. pneumoniae type 4 strain protein from coding region #2299.

XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
XX auditory; respiratory; gene therapy; vaccine.

XX OS Streptococcus pneumoniae type 4 strain.

XX PN WO200277021-A2.

XX PD 03-OCT-2002.

XX PF 27-MAR-2002; 2002WO-IB02163.

XX PR 27-MAR-2001; 2001GB-0007658.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Maignani V, Tettein H, Fraser C;

XX DR WPI, 2003-040579/03.

XX N-PSDB; ABX08011.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
PT or ear infection

PS Claim 1; SEQ ID No 4598; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 693 AA;

Query Match 100.0%; Score 889; DB 24; Length 693;  
Best Local Similarity 100.0%; Pred. No. 2.3e-62;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEARRADAKGKRGVPEGLATPPDKENDASSDSVGEETLPS 60  
DB 265 IKTDREAEAEARRADAKGKRGVPEGLATPPDKENDASSDSVGEETLPS 324  
QY 61 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAEL 120  
DB 325 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLEIAESDVEVKAEL 384  
QY 121 VKEBAKEPRNEEKVQAKAEVESKKAATRLLEKIKTDKKAAEEBAKKAEDVKKEKPA 180  
DB 385 VKEBAKEPRNEEKVQAKAEVESKKAATRLLEKIKTDKKAAEEBAKKAEDVKKEKPA 444

RESULT 13

AAV81653  
ID AAV81653 standard; Protein; 694 AA.

XX AC AAV81653;

XX DT 24-MAY-2000 (first entry)

XX DE Streptococcus pneumoniae protein sequence ID302.

XX KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
XX pneumococcal disease.

XX OS Streptococcus pneumoniae.

XX PN WO200006737-A2.

```
XX 10-FEB-2000.
PD 27-JUL-1999; 99WO-GB02451.
XX 27-JUL-1998; 98GB-0016337.
XX 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX (MICR-) MICROBIAL TECHNICS LTD.
PA
XX Gilbert CFG, Hansbro PM,
PI WPI; 2000-195300/17.
DR
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX
XX Claim 2; Page 96; 108pp; English.
PS
XX AA051501 to AA051679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AA051407 to AA051590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and anti-inflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AA055591 to AA055614 represent primers used in the
CC exemplification of the present invention.
XX
SQ Sequence 694 AA;
Query Match 100.0%; Score 889; DB 21; Length 694;
Best Local Similarity 100.0%; Pred. No. 2.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 265 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPPDKKENDAKSSDSSVGEETLPS 324
QY 61 PSLKPEKKVAEAEKKVEBAKKKADOKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
DB 325 PSLKPEKKVAEAEKKVEBAKKKADOKEDRRNYPNTYKTLLEIAESDVEVKKALEL 384
QY 121 VKEBAKEPRNEKVKQAAEAVESKKAATRLKIKTRDKKAEEBAKRAAEEDVKKEKPA 180
DB 385 VKEBAKEPRNEKVKQAAEAVESKKAATRLKIKTRDKKAEEBAKRAAEEDVKKEKPA 444
RESULT 14
AA032189
ID AA032189 standard; Protein; 460 AA.
XX
XX AA032189;
AC
XX
XX 01-FEB-2000 (first entry)
DT
XX
XX N-terminal choline binding protein A (Cbpa) truncate.
DE
XX Choline binding protein; Cbpa; truncate; adhesin; immunogen;
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KW otitis media; pneumonia.
XX
XX Streptococcus pneumoniae.
OS Synthetic.
OS
XX WO951188-A2.
PN
XX
XX 14-OCT-1999.
PD
```

```
XX 07-APR-1999; 99WO-US07669.
PF
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMUNE INC.
XX
XX Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;
PI WPI; 1999-633690/54.
DR
XX
XX New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections
XX
XX Claim 47; Fig 2A-B; 160pp; English.
PS
XX
XX The present sequence represents an N-terminal choline binding
CC protein A (Cbpa) truncate, denoted Ntype 4 Cbpa trun, derived from
CC Streptococcus pneumoniae Cbpa. Claimed vaccines contain and
CC N-terminal Cbpa truncate such as the present sequence, or a
CC polypeptide comprising a conserved region of the Cbpa truncate.
CC The vaccines provide protection or elicit protective antibodies to
CC bacterial infection, specifically pneumococcus. Antibodies and
CC and antagonists against the N-terminal Cbpa truncates are used in
CC diagnosis and passive immunotherapy.
XX
SQ Sequence 460 AA;
Query Match 99.7%; Score 886; DB 20; Length 460;
Best Local Similarity 99.4%; Pred. No. 2.4e-62;
Matches 179; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 242 IKTDREBAEBAEKRRADAKGQKPKGRARGVPELATPPDKKENDAKSSDSSVGEETLPS 301
QY 61 PSLKPEKKVAEAEKKVEBAKKKADOKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
DB 302 PSLKPEKKVAEAEKKVEBAKKKADOKEDRRNYPNTYKTLLEIAESDVEVKKALEL 361
QY 121 VKEBAKEPRNEKVKQAAEAVESKKAATRLKIKTRDKKAEEBAKRAAEEDVKKEKPA 180
DB 362 VKEBAKEPRNEKVKQAAEAVESKKAATRLKIKTRDKKAEEBAKRAAEEDVKKEKPA 421
RESULT 15
AA049140
ID AA049140 standard; Protein; 446 AA.
XX
XX AA049140;
AC
XX
XX 17-JAN-2000 (first entry)
DT
XX
XX Amino acid sequence of choline-binding protein fragment #4.
DE
XX Truncated surface binding protein; alpha helix; choline binding protein;
KW vaccine; invasive bacterial infection; otitis media; sepsis;
KW meningitis; lobar pneumonia infection; antibody; immature immune system;
KW immunocompromised.
XX
XX Streptococcus pneumoniae.
OS
XX Key Location/Qualifiers
FH Misc-difference 171
FT /note= "Encoded by AAT"
FT Misc-difference 172
FT /note= "Encoded by TAC"
FT Misc-difference 173
FT /note= "Encoded by TTA"
FT Misc-difference 174
```

/note= "Encoded by CAA"

FT  
XX  
XX  
PN WO9951266-A2.  
XX  
XX 14-OCT-1999.  
XX  
XX 06-APR-1999; 99WO-US07680.  
XX  
XX 07-APR-1998; 98US-0080878.  
PR 15-MAY-1998; 98US-0085743.  
XX  
XX (MEDI-) MEDIMUNE INC.  
XX  
XX  
XX Wizemann TM, Koenig S, Johnson LS;  
XX  
XX WPI: 1999-601465/51.  
DR N-PSDB; AA231080.  
XX  
XX  
PT New pneumococcal proteins useful as vaccines and for diagnosis of  
PT pneumococcal infections -  
XX  
XX  
XX Claim 10; Page 68-69; 98pp; English.  
XX  
XX  
CC AAY49137-Y49152 are amino acid sequences that are fragments of choline  
CC binding proteins (CBP). The fragments of the protein are the alpha helix  
CC forming parts of the CBPs from Streptococcus pneumoniae. The  
CC polypeptides do not contain the actual choline binding fragment. The  
CC polypeptides and the nucleotide sequences that encode them  
CC (AA231077-231092) are used in the invention, which relates to polypeptide  
CC truncates of a pneumococcal surface binding protein containing the highly  
CC conserved immunogenic alpha helical portion and no choline binding  
CC portion. The polypeptides are used as immunogens in a bacterial vaccine.  
CC The vaccine can be used for preventing (immunising) or treating invasive  
CC bacterial (especially pneumococcal) infections, especially otitis media  
CC (caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia  
CC infections. Antibodies raised against the polypeptide are useful for  
CC detection, prevention (passive immunity) and treatment of S. pneumoniae  
CC infections. The vaccines are especially useful in immunocompromised  
CC patients, those with an immature immune system, or patients with an on  
CC going pneumococcal infection. The vaccine avoids unnecessary expense and  
CC provides broad protection against a range of pneumococcal serotypes and  
CC it produces an improved and enhanced effect in preventing bacterial  
CC infections.  
CC  
SQ Sequence 446 AA;

Query Match 99.4%; Score 884; DB 20; Length 446;  
Best Local Similarity 99.4%; Pred. No. 3.4e-62;  
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IKTDREAEAEARRADAKGQKPKRAKGVPGELATPDKKENDAKSSDSSVGEETLPS 60  
DB 228 IKTDREAEAEARRADAKGQKPKRPGVPGELATPDKKENDAKSSDSSVGEETLPS 287  
QY 61 PSLRPEKVAEAEKVEAEKKAEDQKEDRRNYPNTNTYKTLLELEIASDVEVKKAEL 120  
DB 288 PSLRPEKVAEAEKVEAEKKAEDQKEDRRNYPNTNTYKTLLELEIASDVEVKKAEL 347  
QY 121 VKEEAKEPRNEKVKQAKAEVSKKAEATRLKIKTDREKAEAEKRAAEEDVKKEKPA 180  
DB 348 VKEEAKEPRNEKVKQAKAEVSKKAEATRLKIKTDREKAEAEKRAAEEDVKKEKPA 407

Search completed: November 21, 2003, 13:31:58  
Job time : 19.5369 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:28:52 ; Search time 47.3285 Seconds  
(without alignments)  
2786.162 Million cell updates/sec

Title: US-09-298-523C-13  
Perfect score: 2537  
Sequence: 1 MFASKSRKHYHSIRKFSVG.....NRLTQQQPKTEKPAQDPSTP 511

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	16 Q97N74	Q97N74 streptococc
2	2507	98.8	693	2 Q9KX21	Q9KX21 streptococc
3	2473.5	97.5	696	2 Q9KX32	Q9KX32 streptococc
4	2459.5	96.9	516	2 Q9R0T2	Q9R0T2 streptococc
5	2390	94.2	695	2 Q9KX36	Q9KX36 streptococc
6	2040.5	80.4	684	2 Q9KX17	Q9KX17 streptococc
7	1788	70.5	695	2 Q9KX51	Q9KX51 streptococc
8	1782	70.2	681	2 Q9KX16	Q9KX16 streptococc
9	1780.5	70.2	660	2 Q9KX18	Q9KX18 streptococc
10	1773	69.9	709	2 Q9KX38	Q9KX38 streptococc
11	1770.5	69.8	487	2 Q9R0T3	Q9R0T3 streptococc
12	1754.5	69.2	752	2 Q9KX30	Q9KX30 streptococc
13	1732.5	68.3	752	2 Q9KX33	Q9KX33 streptococc
14	1730	68.2	564	2 Q69188	Q69188 streptococc
15	1730	68.2	701	2 Q9KX48	Q9KX48 streptococc
16	1730	68.2	701	16 Q8DN05	Q8DN05 streptococc

17	1726	68.0	701	2 Q9R0T5	Q9R0T5 streptococc
18	1725.5	68.0	730	2 Q9KX47	Q9KX47 streptococc
19	1718	67.7	581	2 Q33742	Q33742 streptococc
20	1715.5	67.6	660	2 Q9KX41	Q9KX41 streptococc
21	1713.5	67.5	680	2 Q9KX43	Q9KX43 streptococc
22	1713.5	67.5	730	2 Q9KX31	Q9KX31 streptococc
23	1689	66.6	699	2 Q9KX42	Q9KX42 streptococc
24	1683	66.3	655	2 Q9KX50	Q9KX50 streptococc
25	1631	64.3	657	2 Q9KX29	Q9KX29 streptococc
26	1616	63.7	684	2 Q9KX46	Q9KX46 streptococc
27	1609.5	63.4	488	2 Q9R0T6	Q9R0T6 streptococc
28	1591.5	62.7	678	2 Q9KX49	Q9KX49 streptococc
29	1589.5	62.7	678	2 Q9KX49	Q9KX49 streptococc
30	1582.5	62.4	680	2 Q9KX11	Q9KX11 streptococc
31	1578	62.2	681	2 Q9KX13	Q9KX13 streptococc
32	1573	62.0	676	2 Q9KX39	Q9KX39 streptococc
33	1573	62.0	666	2 Q9KX14	Q9KX14 streptococc
34	1569.5	61.9	663	2 Q30874	Q30874 streptococc
35	1398	55.1	739	2 Q9R0T4	Q9R0T4 streptococc
36	1345	53.0	820	2 Q9R0T1	Q9R0T1 streptococc
37	1343.5	53.0	869	2 Q9KX27	Q9KX27 streptococc
38	1330	52.4	929	2 Q9ZAY5	Q9ZAY5 streptococc
39	1330	52.4	929	2 Q9KX19	Q9KX19 streptococc
40	1225	48.3	565	2 Q9KX15	Q9KX15 streptococc
41	1198.5	47.2	769	2 Q9KX40	Q9KX40 streptococc
42	1195	47.1	667	2 Q9KX28	Q9KX28 streptococc
43	1175	46.3	523	2 Q33753	Q33753 streptococc
44	1165.5	45.9	770	2 Q9KX37	Q9KX37 streptococc
45	1161.5	45.8	769	2 Q9PD01	Q9PD01 streptococc

#### ALIGNMENTS

##### RESULT 1

ID Q97N74 PRELIMINARY: PRT; 693 AA.

AC Q97N74; 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Choline binding protein A.

GN SP2190.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TIGR4;

RX MEDLINE=21357209; PubMed=11463916;

RA Tetteilin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,

RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,

RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,

RA Holtzapfel E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,

RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,

RA Holt I.E., Loftus B.D., Yang F., Smith H.O., Venter J.C.,

RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;

RT "Complete genome sequence of a virulent isolate of Streptococcus

pneumoniae";

RL Science 293:498-506(2001).

DR EMBL: AE007507; AAK76241.1; -.

DR TIGR: SP2190; -.

DR InterPro: IPR0052479; CW binding.

DR Pfam: PF04650; YSIRK\_signal\_1; 8.

DR TIGRPFAM: TIGR01168; YSIRK\_signal\_1; 1.

DR Complete proteome.

SQ SEQUENCE 693 AA; 77762 MW; 6F8F47C32E344A41 CRC64;

Query Match 99.2%; Score 2516; DB 16; Length 693;



```
Db 61 GEOPKLLDSEDRKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNINIKNEYLNTIV 120
Qy 120 ESTSESOLQILMMESRSKYDEAVSKFEKDSSSSSSSSTKPEASDTAKPKNPTPEGKV 179
Db 121 ESTSESOLQILMMESRSKYDEAVSKFEKDSSSSSSSSTKPEASDTAKPKNPTPEGKV 180
Qy 180 AEAKKKVEEA-KKADQKEEDRRNYPITTYKTLELAIASDVEYKKALELVKVKANEP 238
Db 181 AEAKKKVEEA-KKADQKEEDRRNYPITTYKTLELAIASDVEYKKALELVKVKANEP 240
Qy 239 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARADAKQGPRAKAGV 294
Db 241 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARADAKQGPRAKAGV 300
Qy 295 PGEIATPKKNDKSSSSSGEETLPSPSLKPEKVAEAKKVEAKKADQKEEDRR 354
Db 301 PGEIATPKKNDKSSSSSGEETLPSPSLKPEKVAEAKKVEAKKADQKEEDRR 360
Qy 355 NYPTTYKTLELAIASDVEYKKALELVKEAKEPRNEEKVKQAKAEVESKKAATRL 414
Db 361 NYPTTYKTLELAIASDVEYKKALELVKEAKEPRNEEKVKQAKAEVESKKAATRL 420
Qy 415 KIKTDKKAEEBAKKAEDKVKKEPAEQOPAPAPAKEPAPAKPENPAEQPKAEXP 474
Db 421 KIKTDKKAEEBAKKAEDKVKKEPAEQOPAPAPAKEPAPAKPENPAEQPKAEXP 480
Qy 475 ADQAEEDYARRSEEVNRLTQOOPKTEKPAOPSTP 511
Db 481 ADQAEEDYARRSEEVNRLTQOOPKTEKPAOPSTP 516
```

## RESULT 4

```
Q9ROT2 PRELIMINARY; PRT; 516 AA.
ID Q9ROT2
AC Q9ROT2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 57.9 kDa protein (Fragment).
GN pSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L81905;
RX MEDLINE=20038319; PubMed=10569772;
RA Brooks-Walter A., Briles D.E., Hollingshead S.K.;
RT "The pSpC gene of streptococcus pneumoniae encodes a polymorphic
protein, pSpC, which elicits cross-reactive antibodies to Pepsa and
RT provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67:6533-6542 (1999).
DR EMBL; AF068649; AAF13459.1; -
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
DR Hypothetical protein.
KW NON TER
FT 516
SQ SEQUENCE 516 AA; 57886 MW; 1119782688BB7E35 CRC64;
```

```
Query Match 96.9%; Score 2459.5; DB 2; Length 516;
Best Local Similarity 97.9%; Pred. No. 2.8e-99;
Matches 502; Conservative 2; Mismatches 6; Indels 3; Gaps 3;
```

Qy 1 MFASKSEBRKYHSIRKFSVG-ASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59  
Db 1 MFASKSEBRKYHSIRKFSVGVSFVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59

Qy 60 GEOPKLLDSEDRKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNINIKNEYLNTIV 119  
Db 60 GEOPKLLDSEDRKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNINIKNEYLNTIV 119

## RESULT 5

```
Q9KK36 PRELIMINARY; PRT; 695 AA.
ID Q9KK36
AC Q9KK36;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN pSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g394;
RA Ianneli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154023; AAF73792.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 695 AA; 78361 MW; F8C84F08C4EFPFB4 CRC64;
```

Query Match 94.2%; Score 2390; DB 2; Length 695;  
Best Local Similarity 95.2%; Pred. No. 3.7e-96;  
Matches 492; Conservative 6; Mismatches 11; Indels 8; Gaps 4;

Qy 1 MFASKSEBRKYHSIRKFSVG-ASVVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 59  
Db 1 MFASKSEBRKYHSIRKFSIGVASVVASLVMSGVVHATENEGATQVPTSSNRANESQAEQ 60

Qy 60 GEOPKLLDSEDRKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNINIKNEYLNTIV 119  
Db 61 GEOPKLLDSEDRKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNINIKNEYLNTIV 120

Qy 120 ESTSESOLQILMMESRSKYDEAVSKFEKDSSSSSSSSTKPEASDTAKPKNPTPEGKV 179  
Db 121 ESTSESOLQILMMESRSKYDEAVSKFEKDSSSSSSSSTKPEASDTAKPKNPTPEGKV 180

```
QY 180 AEAKKKVEEA--KKAQDQKEEDRNYPITYKTLELEIAESDVEYKKALELVKVKANPR 238
DB 181 ABAKKKVEEAEEAKKAQDQKEEDRNYPITYKTLELEIAESDVEYKKALELVKVKANPR 240
QY 239 DEOKIKQAEAEVESKQAEATRLKKIKTDRBEAEBAEKARADAKEOG---KPKGRAGV 294
DB 241 DEBKIKQAEAEVESKQAEATRLKKIKTDRBEAKGEGAKRADAKEDDESKKRSKVRKGD 300
QY 295 PGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAAEDQKEEDRR 354
DB 301 LGEQATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAAEDQKEEDRR 360
QY 355 NYPNTYTTLELEIAESDVEYKKALELVKEBAKPRNEEKVKQAKAEVESKKAATLE 414
DB 361 NYPNTYTTLELEIAESDVEYKKALELVKEBAKPRNEEKVKQAKAEVESKKAATLE 420
QY 415 KIKTRKKAEEBAKKAEEBKVEKEPAEQOPAPAPAEKPAPEKPEENPAEQPAEK 474
DB 421 KIKTRKKAEEBAKKAEEBKVEKEPAEQOPAPAPAEKPAPEKPEENPAEQPAEK 478
QY 475 ADQQAEDYARSEEEYNRLTQQQPPKTEKPAQPS 511
DB 479 ADQQAEDYARSEEEYNRLTQQQPPKTEKPAQPS 515

RESULT 6
Q9KK17 PRELIMINARY; PRT; 684 AA.
ID Q9KK17
AC Q9KK17
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=stre22;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154039; AAF73811.1;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR Pfam; PF01473; CW binding_1; 7.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 684 AA; 77589 MW; 20FB892F82FDDF25 CRC64;

Query Match 80.4%; Score 2040.5; DB 2; Length 684;
Best Local Similarity 80.6%; Pred. No. 4.7e-81;
Matches 425; Conservative 34; Mismatches 49; Indels 19; Gaps 6;
```

```
QY 238 RDEKIKQAEAEVESKQAEATRLKKIKTDRBEAEBAK-----RRADAKGQPKGRAK 291
DB 238 RDEKIKQAEAEVESKQAEATRLKKIKTDRBEAEBAKLEAVERKNAATSSQGPKRVRK 297
QY 292 RGVPEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAAEDQKEE 351
DB 298 RGVPEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAAEDQKEE 357
QY 352 DRRNYPNTYTTLELEIAESDVEYKKALELVKEBAKPRNEEKVKQAKAEVESKKAAT 411
DB 358 DRRNYPNTYTTLELEIAESDVEYKKALELVKEBAKPRNEEKVKQAKAEVESKKAAT 417
QY 412 RLEKIKTRKKAEEBAKKAEEBKVEKEPAEQOPAPAPAEK-----PAPAPKEN 464
DB 418 RLEKIKTRKKAEEBAKKAEEBKVEKEPAEQOPAPAPAEKPEPEENPAEQPAEK 477
QY 465 PAEQPAEKPAEQQAEDYARSEEEYNRLTQQQPPKTEKPAQPS 511
DB 478 PAEQPAEKPAEQQAEDYARSEEEYNRLTQQQPPKTEKPAQPS 524

RESULT 7
Q9KK51 PRELIMINARY; PRT; 695 AA.
ID Q9KK51
AC Q9KK51
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC6307;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154009; AAF73776.1;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR Pfam; PF01473; CW binding_1; 10.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 695 AA; 78521 MW; F208BC7285DA4FCE CRC64;

Query Match 70.5%; Score 1788; DB 2; Length 695;
Best Local Similarity 73.6%; Pred. No. 3.9e-70;
Matches 380; Conservative 39; Mismatches 51; Indels 46; Gaps 7;
```



QY 236 GELATPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAKKEPRNEEKVQAKAEVSK 355  
DB 260 GEPATPDKKENDAKSSDSSVGEETLPSPSLKPGKVAEAKKEPRNEEKVQAKAEVSK 319  
QY 336 YPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSKAEATRLK 415  
DB 320 YPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSKAEATRLK 379  
QY 416 IKTDRKKAEKKEPRNEEKVQAKAELELVKEAKKEPRNEEKVQAKAEVSKAEATRLK 475  
DB 380 IKTDRKKAEKKEPRNEEKVQAKAELELVKEAKKEPRNEEKVQAKAEVSKAEATRLK 439  
QY 476 DQAEEDYARRESEENRNLTOOQPKTEKPAQSTP 511  
DB 440 DQAEEDYARRESEENRNLTOOQPKTEKPAQSTP 475

RESULT 8  
QYK16 PRELIMINARY; PRT; 681 AA.  
AC Q9KK16: 09KK16: 09KK16:  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Surface protein pspc.  
GN pspc.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=stf25;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspc of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF154040; AAF73812.1; -  
DR InterPro: IPR002479; CM binding.  
DR InterPro: IPR005877; Gpos\_YsIRK.  
DR Pfam: PF04650; YsIRK signal; 1.  
DR TIGRFAMs: TIGR01168; YsIRK signal; 1.  
SQ SEQUENCE 681 AA; 76728 MW; 38FE4782653D51A8 CRC64;

Query Match 70.2%; Score 1782; DB 2; Length 681;  
Best Local Similarity 70.6%; Pred. No. 6.9e-70;  
Matches 379; Conservative 43; Mismatches 51; Indels 64; Gaps 9;

QY 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGCATOVPTSSNRANESQAEQ 59  
DB 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGCATOVPTSSNRANESQAEQ 57  
QY 60 GEQPKLDSERDARKEVEEYVKKIVGESYAKSTKKRHTTTVALVNLNINIKNEYLNKIV 119  
DB 58 -----TEYMAAKAQVDEYITKKL-----QLDRRHIONVGLITKLGVITKEYLHRLS 104  
QY 120 ESTSESOQLIMMESRSKVDEAVSKFEKXSSSSSSSSSTKPEASDTAKNKPTEPEKV 179  
DB 105 VSKKESBAE-LPSEIVAKLDAFAEQFKDT-----LPTEPKKV 142  
QY 180 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 238  
DB 143 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 202  
QY 239 DEQKIKOAEVSKOAEATRLKKITDRBEAEKRAADAK-----EQGKPR 266  
DB 203 DEQKIKOAEVSKOAEATRLKKITDRBEAEKRAADAK-----EQGKPR 262  
QY 287 KGRKRGVPELATPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAKKEPRNEEKVQAKAE 346  
DB 263 KGRKRGVPELATPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAKKEPRNEEKVQAKAE 322

QY 347 DQKEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSK 406  
DB 323 DQKEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSK 382  
QY 407 KAEATRLKKITDRKKAEEKRAAEEDVKKEKPAQOPAPAKAEK-----PAPA 459  
DB 383 KAEATRLKKITDRKKAEEKRAAEEDVKKEKPAQOPAPAKAEK-----PAPA 442  
QY 460 PKPENPAQOPAEKPAQOAEEDYARRESEENRNLTOOQPKTEKPA-----QSTP 511  
DB 443 PKPENPAQOPAEKPAQOAEEDYARRESEENRNLTOOQPKTEKPA-----QSTP 499

RESULT 9  
QYK18 PRELIMINARY; PRT; 660 AA.  
AC Q9KK18: 09KK18: 09KK18:  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Surface protein pspc.  
GN pspc.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=stf2;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspc of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF154038; AAF73810.1; -  
DR InterPro: IPR002479; CM binding.  
DR InterPro: IPR005877; Gpos\_YsIRK.  
DR Pfam: PF04650; YsIRK signal; 1.  
DR TIGRFAMs: TIGR01168; YsIRK signal; 1.  
SQ SEQUENCE 660 AA; 74515 MW; 374EB582FD09A659 CRC64;

Query Match 70.2%; Score 1780.5; DB 2; Length 660;  
Best Local Similarity 73.4%; Pred. No. 7.8e-70;  
Matches 383; Conservative 38; Mismatches 48; Indels 53; Gaps 9;

QY 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGCATOVPTSSNRANESQAEQ 59  
DB 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGCATOVPTSSNRANESQAEQ 57  
QY 60 GEQPKLDSERDARKEVEEYVKKIVGESYAKSTKKRHTTTVALVNLNINIKNEYLNKIV 119  
DB 58 -----TEYMAAKAQVDEYITKKL-----QLDRRHIONVGLITKLGVITKEYLHRLS 104  
QY 120 ESTSESOQLIMMESRSKVDEAVSKFEKXSSSSSSSSSTKPEASDTAKNKPTEPEKV 179  
DB 105 VSKKESBAE-LPSEIVAKLDAFAEQFKDT-----LPTEPKKV 142  
QY 180 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 238  
DB 143 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 202  
QY 239 DEQKIKOAEVSKOAEATRLKKITDRBEAEKRAADAK-----EQGKPR 289  
DB 203 DEQKIKOAEVSKOAEATRLKKITDRBEAEKRAADAK-----EQGKPR 261  
QY 290 AKRGVGEATLPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAKKEPRNEEKVQAKAE 349  
DB 262 ANRGVGEATLPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAKKEPRNEEKVQAKAE 321  
QY 350 EEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSKAE 409  
DB 322 EEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKKEPRNEEKVQAKAEVSKAE 381

OY 410 ATRLEKITDPRKKAEEBAKKAEEADKYKEKPAEOPAPAPAKPAAPAPKPAOP 469  
DB 362 ATRLEKITDPRKKAEEBAKKAEEADKYKEKPAEOPAPAKPAOP 438  
OY 470 KAEKPADOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 511  
DB 439 KAEKPADOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 480

RESULT 10  
OQK38 PRELIMINARY; PRT; 709 AA.  
AC OQK38;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PSpC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=g387;  
RA Iannelli F., Ovgioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF154021; AAF7390.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR InterPro; IPR00533; Troponyosin.  
DR Pfam; PF01473; CW\_binding\_1; 9.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR PRINTS; PR00194; TROPOMYOSIN.  
DR TIGRfams; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 709 AA; 80251 MW; 32BBC96E380EBB7A CRC64;

Query Match 69.9%; Score 1773; DB 2; Length 709;  
Best Local Similarity 70.8%; Pred. No. 1.6e-69;  
Matches 380; Conservative 49; Mismatches 54; Indels 54; Gaps 11;  
OY 1 MFASSEKRYHYSTRKESVG-ASVVVASLWGSVVAH-ENEGATQVTSNRANESQ-- 56  
DB 1 MFASNEERKHYSTRKESVGASVAVASLFGSVVAHAEAGNYPSSSCSSSEAKKSQTE 60  
OY 57 -----AEOGEPKLDSEBDKARKEVEEYVKIVGESYAKSTKKRTITVALVNLNINIKN 112  
DB 61 HMKAAEQCKHREIDLDRNKAKAIDEYIEKMLSE--IQDGRKKTQNFALNKLRIKTIKT 118  
OY 113 EYLN--KIVESTSESQLOI---LMMESRSKYDEAVSKFEKSSSSSSSSSTKPEASDTA 167  
DB 119 EYLVGLSVLKESESELPKAKELPEKIKELKLAEPHFKKDT----- 160  
OY 168 KPNKPTBEGEKYAEAKKKVVEA-KKAKOQKEEDRRNYPTITTYKTLELAESDVEYKAE 226  
DB 161 -----LRPEGEYAEAKKKVVEAEEKKAKOQKEDYENYPTITTYKTLELAESDVEYKAE 215  
OY 227 LEIVVKNBPRDEKIKQAEAEVESKQAEATRLKKITDREBEAEKRRADAK----- 281  
DB 216 LEIVVEANKPRNEQKVQAKAVESKQAEATRLKKITDREBEAEKRRADAKAV 275  
OY 282 -----EQGPKRAGVPGELATPDKENDAKSSDSVGEETLPSPSLKPEKVAEA 334  
DB 276 EKNATTSQGPKRAGVPGELATPDKENDAKSSDSVGEETLPSPSLKPEKVAEA 335  
OY 335 EKQVEAEAKKADQKEEDRRNYPTITTYKTLELAESVVEYKAELEIVKEAEKPRNEE 394  
DB 336 EKQVEAEAKKADQKEEDRRNYPTITTYKTLELAESVVEYKAELEIVKEAEKPRDEE 395  
OY 395 KVKQAKAEVESKKAATLEKIKITDRKKAEEBAKKAEEEDKYKEKPAEOPAPAPAE 454

DB 396 KIKQAKAEVESKKAATLEKIKITDRKKA--EBAKKAEEEDKYKEKPAEOPAPAPQE 454  
OY 455 KPAAPKPRNPAPQPKAEKPADQOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 511  
DB 455 K--PAPKPEKPAEHPKAKENRADQOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 509

RESULT 11  
OQROT3 PRELIMINARY; PRT; 487 AA.  
AC OQROT3;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE Hypothetical 55.0 kDa protein (Fragment).  
GN PSpC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bg8090;  
RX MEDLINE=20038319; PubMed=10569772;  
RA Brooks-Walter A., Briles D.E., Hollinghead S.K.;  
RT "The pspC gene of streptococcus pneumoniae encodes a polymorphic  
RT protein, PspC, which elicits cross-reactive antibodies to PspA and  
RT provides immunity to pneumococcal bacteremia."  
RL Infect. Immun. 67:6533-6542(1999).  
DR EMBL; AF068648; AAF13458.1; -  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRfams; TIGR01168; YsIRK\_signal; 1.  
DR Hypothetical protein.  
FT NON TER 487  
SQ SEQUENCE 487 AA; 54962 MW; 7BDBB21C65341E86 CRC64;

Query Match 69.8%; Score 1770.5; DB 2; Length 487;  
Best Local Similarity 71.9%; Pred. No. 1.6e-69;  
Matches 378; Conservative 43; Mismatches 50; Indels 55; Gaps 8;  
OY 1 MFASSEKRYHYSTRKESVG-ASVVVASLWGSVVAH-ENEGATQVTSNRANESQAEQ 59  
DB 1 MFASNEERKHYSTRKESVGASVAVASLFGSVVAHTEKEVTTQVATSPKAKKSQ--- 57  
OY 60 GEOPKLDSEBDKARKEVEEYVKIVGESYAKSTKKRTITVALVNLNINIKNEYLNKIV 119  
DB 58 -----TEHMKAAKQVDEYITKTL-----QLDRRKHQVGLTKLVITKTEYLRHS 104  
OY 120 ESTSESQLOILMMSRSKVDVAVKFEKSSSSSSSSSTKPEASDTAKPRKPTBEGEKV 179  
DB 105 VSKESSEAE-LPSIKAKLDAAEQFKKDT-----LPTPEPKV 142  
OY 180 AEAKKYVEA-KKADQKEEDRRNYPTITTYKTLELAESDVEYKAELEIVKANEPR 238  
DB 143 AEAKKYVEAKKADQKEEDRRNYPTITTYKTLELAESDVEYKAELEIVKEAAGSR 202  
OY 239 DEQIKQAEAEVESKQAEATRLKKIKITDREAE-----EBAKRRADAEQG 284  
DB 203 NEQVKNQAKKAVESKQAEATRLKKIKITDREAEATRLLENITKDRKAEBAKRRADAEQD 262  
OY 285 KPKRAGVPGELATPDKENDAKSSDSVGEETLPSPSLKPEKVAEAEAKKYAEAKK 344  
DB 263 ESKRVRGVGPGEQATLDKENDAKSSDSVGEETLPSPSLKPEKVAEAEAKKYAEAKK 322  
OY 345 AEQDKEEDRRNYPTITTYKTLELAESDVEYKAELEIVKEAEKPRNEEKVQAKAEVE 404  
DB 323 AKQDKEEDRRNYPTITTYKTLELAESDVEYKAELEIVKEAEKPRNEEKVQAKAEVE 382  
OY 405 SKKAEATRLKKITDRKKAEEBAKKAEEADKYKEKPAEOPAPAPAKPAAPAPPEN 464  
DB 383 SKKAEATRLKKITDRKKA--EBAKKAEEADKYKEKPAEOPAPAPAPAPAPAPPEN 441

Qy 465 PAEPKAEKPADQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 510  
 Db 442 PAEPKAEKPADQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 487

RESULT 12

Q9KK30 PRELIMINARY; PRT; 752 AA.  
 ID 09KK30  
 AC 09KK30  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Surface protein PspC.  
 GN PspC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_Taxid=1313;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=g408;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF154029; AAF73798.1;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR005877; Gpos\_YsIRK.  
 DR Pfam: PF01473; CW\_binding\_1; 12.  
 DR Pfam: PF04650; YsIRK\_signal; 1.  
 DR TIGRFAMs: TIGR01168; YsIRK\_signal; 1.  
 SQ SEQUENCE 752 AA; 85296 MW; D350BBBCFC2ABD3 CRC64;

Query Match 69.2%; Score 1754.5; DB 2; Length 752;  
 Best Local Similarity 70.8%; Pred. No. 1.2e-68;  
 Matches 379; Conservative 41; Mismatches 48; Indels 67; Gaps 11;

Qy 1 MFASKSRKHYHSIRKRSVG-ASVVVASLWGSVVAHATENEGATQVPTSSNRANESQAEQ 59  
 Db 1 MFASKSRKHYHSIRKRSVGASVVVASLFGVVAH---EGVRSWDT----- 45  
 Qy 60 GEOPKLDSEKDKARK---EVEEYVKKIIGESYAKSTKRRHTITVALVNLNKKOEYLNK 115  
 Db 46 ---PKVTPSGQNTSKKAPDEVKSHLEKILSE--IQDKRKHTONLANKLSRIQTEYFY 100  
 Qy 116 ---NKIVESTSE---SOLQIMESRSKVDKAVSKFEKSSSSSSSSSTKPEASDT 166  
 Db 101 LKKKLKALSLTSKTEBELTSKTEBELTSKTELDAAFGFKDITLS----- 145  
 Qy 167 AKPNKPTPEQEKVAEAKKQVEEA-KKAKDQKEEDRRNYPTITYKTLELEIAESDVEYKA 225  
 Db 146 ---TEPEKVAEAKQKVAEAKKADQKEEDRRNYPTITYKTLELEIAEFVAKYKA 139  
 Qy 226 ELETVKVAEAPREKIKIOAEAVESKQAEATRLKKIKTDREAEAEARADAKE--- 282  
 Db 200 ELKLKAKAOTRDEEIKIOAEAVESKQAEATRLKKIKTDREAEAEARADAKE 259  
 Qy 283 ---QGPKGRAGRVGPELATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAK 336  
 Db 260 EAPASDQGRKGRAGRVGPELATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAK 319  
 Qy 337 KVEBAKKAEDQKEEDRRNYPTITYKTLELEIAESDVEYKKALELVKEAEKPRNEKY 396  
 Db 320 KVEBAKKAEDQKEEDRRNYPTITYKTLELEIAESDVEYKKALELVKEAEKPRNEKY 379  
 Qy 397 KQAEAVESKKAATRELEKIKTDREAEAEARADAKE--- 456  
 Db 380 KQAEAVESKKAATRELEKIKTDREAEAEARADAKE--- 438  
 Qy 457 APAKPEKPAQAEKPADQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 511  
 Db 439 -PAKPEKPAQAEKPADQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 492

RESULT 13

Q9KK33 PRELIMINARY; PRT; 752 AA.  
 ID 09KK33  
 AC 09KK33  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Surface protein PspC.  
 GN PspC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_Taxid=1313;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=g408;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
 RT Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF154026; AAF73795.1;  
 DR InterPro: IPR002479; CW\_binding.  
 DR InterPro: IPR005877; Gpos\_YsIRK.  
 DR Pfam: PF01473; CW\_binding\_1; 13.  
 DR Pfam: PF04650; YsIRK\_signal; 1.  
 DR TIGRFAMs: TIGR01168; YsIRK\_signal; 1.  
 SQ SEQUENCE 752 AA; 85321 MW; E24BF1DBF50A7F8 CRC64;

Query Match 68.3%; Score 1732.5; DB 2; Length 752;  
 Best Local Similarity 71.1%; Pred. No. 1e-67;  
 Matches 366; Conservative 55; Mismatches 47; Indels 47; Gaps 8;

Qy 1 MFASKSRKHYHSIRKRSVG-ASVVVASLWGSVVAHATENEGATQVPTSSNRANESQAEQ 59  
 Db 1 MFASKSRKHYHSIRKRSVGASVVVASLFGVVAH---EGVRSWDT----- 45  
 Qy 60 GEOPKLDSEKDKARK---EVEEYVKKIIGESYAKSTKRRHTITVALVNLNKKOEYLNK 119  
 Db 58 ---TEHMKARQVDEYVANKMI---QDKRKHTONLANKLSRIQTEYFY 104  
 Qy 120 ESTSESOLQIMESRSKVDKAVSKFEKSSSSSSSSSTKPEASDTAKNKPTEPEKY 179  
 Db 105 VLEKKRKEBELTSKTEBELTSKTEBELTSKTELDAAFGFKDITLS----- 141  
 Qy 180 AEAKKQVEEA-KKAKDQKEEDRRNYPTITYKTLELEIAESDVEYKKALELVKVAEAPR 238  
 Db 142 EEOKKVEEAEKKAQKEEDRRNYPTITYKTLELEIAESDVEYKKALELVKVAEAPR 201  
 Qy 239 DEOKIKQAEAVESKQAEATRLKKIKTDREAEAEARADAKEQ--GPKGRAGRVG 296  
 Db 202 NEBKIKKAEAVESKQAEATRLKKIKTDREAEAEARADAKEQ--GPKGRAGRVG 261  
 Qy 297 ELATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAKKADQKEEDRRNY 356  
 Db 262 EAPATPDKKENDAKSSDSSVGEETLPSLSLPEKKVAEAKKADQKEEDRRNY 321  
 Qy 357 PTITYKTLELEIAESDVEYKKALELVKEAEKPRNEKYKQAEAVESKKAATRELEK 416  
 Db 322 PTITYKTLELEIAESDVEYKKALELVKEAEKPRNEKYKQAEAVESKKAATRELEK 381  
 Qy 417 KTDREAEAEARADAKE--- 476  
 Db 382 KTDREAEAEARADAKE--- 437  
 Qy 477 QQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 511  
 Db 438 QQAEEDYARSEEEYNRLTQOQPPKTEKPAQPSST 472

RESULT 14

069186

ID 069188 PRELIMINARY; PRT; 564 AA.  
AC 069188;  
DT 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE C3-binding protein.  
GN PBCA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OK NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CP1200;  
RX MEDLINE=20281293; PubMed=10820017;  
RA Cheng Q., Finkel D., Hostetter M.K.;  
RT "Novel purification scheme and functions for a C3-binding protein from  
RT Streptococcus pneumoniae."  
RL Biochemistry 39:5450-5457(2000).  
DR EMBL: AF067128; AAC17445.1; -;  
DR InterPro; IPR002479; Gpos\_binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CM\_binding\_1; 4.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 564 AA; 63596 MW; 884C97BAA04AED20 CRC64;

Query Match 68.2%; Score 1730; DB 2; Length 564;  
Best Local Similarity 70.7%; Pred. No. 1e-67;  
Matches 371; Conservative 49; Mismatches 47; Indels 58; Gaps 10;

OY 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEO 59  
DB 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEO 54  
OY 60 GEOPKLDSEBDRKARKE-VEEYVKIVGESYAKSTKKRHTTTVALVNELNINIKNEYLNK- 117  
DB 55 -----TEHRKAKAQVDEYIEKMLRE--IQDRKHQTNVALNITLSAIKTKYLREL 104  
OY 118 -IVESTSSQOILMMESRSKVDKAVSKFEKDSSSSSSDSTKPEASDTAKPNKPTPEG 176  
DB 105 NVLEEKSKDELPL--SEIKAKLDAFAFEKFKKDT-----LKPFG 138  
OY 177 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVKAN 235  
DB 139 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAK 198  
OY 236 EPRDEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAK-----EOKGP 286  
DB 199 ESRNEGITIKQAEVSKQAEATRLLENIKTDREKAEAEAKRADAKLEANVATSDQCKP 258  
OY 287 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVAAE 346  
DB 259 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVAAE 318  
OY 347 DQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAEKPEPRNEKVKQAAVESEK 406  
DB 319 DQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAEKPEPRNEKVKQAAVESEK 378  
OY 407 KAEATRLLEIKITDRKKAEEAEKRAAEEDKVEKPAEPAPAPAKPEKPAAPKPEKPA 466  
DB 379 KAEATRLLEIKITDRKKAEEAEKRAAEEDKVEKPAEPAPAPAKPEKPAAPKPEKPA 436  
OY 467 EOPKAEKPADQQAEDYARRSEEEYNRLTQOOPKTEKPAQPSPTP 511  
DB 437 EOPKAEKPADQQAEDYARRSEEEYNRLTQOOPKTEKPAQPSPTP 481

RESULT 15  
OYK48 PRELIMINARY; PRT; 701 AA.  
AC 09KK48;  
DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE Surface protein pspC.  
GN PSPC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OK NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D39;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154012; AAF73779.1; -;  
DR InterPro; IPR002479; Gpos\_binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CM\_binding\_1; 10.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 701 AA; 79098 MW; A80E31FE2846F1BF CRC64;

Query Match 68.2%; Score 1730; DB 2; Length 701;  
Best Local Similarity 70.7%; Pred. No. 1.3e-67;  
Matches 371; Conservative 49; Mismatches 47; Indels 58; Gaps 10;

OY 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEO 59  
DB 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEO 54  
OY 60 GEOPKLDSEBDRKARKE-VEEYVKIVGESYAKSTKKRHTTTVALVNELNINIKNEYLNK- 117  
DB 55 -----TEHRKAKAQVDEYIEKMLRE--IQDRKHQTNVALNITLSAIKTKYLREL 104  
OY 118 -IVESTSSQOILMMESRSKVDKAVSKFEKDSSSSSSDSTKPEASDTAKPNKPTPEG 176  
DB 105 NVLEEKSKDELPL--SEIKAKLDAFAFEKFKKDT-----LKPFG 138  
OY 177 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVKAN 235  
DB 139 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAK 198  
OY 236 EPRDEQIKQAEAEVSKQAEATRLKKIKTDREAEAEAKRADAK-----EOKGP 286  
DB 199 ESRNEGITIKQAEVSKQAEATRLLENIKTDREKAEAEAKRADAKLEANVATSDQCKP 258  
OY 287 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVAAE 346  
DB 259 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVAAE 318  
OY 347 DQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAEKPEPRNEKVKQAAVESEK 406  
DB 319 DQKEEDRRNYPITTYKTLELEIAESDVVKAELELVKVEAEKPEPRNEKVKQAAVESEK 378  
OY 407 KAEATRLLEIKITDRKKAEEAEKRAAEEDKVEKPAEPAPAPAKPEKPAAPKPEKPA 466  
DB 379 KAEATRLLEIKITDRKKAEEAEKRAAEEDKVEKPAEPAPAPAKPEKPAAPKPEKPA 436  
OY 467 EOPKAEKPADQQAEDYARRSEEEYNRLTQOOPKTEKPAQPSPTP 511  
DB 437 EOPKAEKPADQQAEDYARRSEEEYNRLTQOOPKTEKPAQPSPTP 481

Search completed: November 21, 2003, 13:33:49  
Job time : 51.3285 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:29:32 ; Search time 7.29378 Seconds  
(without alignment)  
2373.306 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442  
Perfect score: 889  
Sequence: 1 IKTDREBEAEAKRRADAKE.....AEEAKRKAEDKVEKPA 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR.76:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	693	2 H95255	choiline binding pr
2	803.5	90.4	701	2 H98120	choiline binding pr
3	194	21.8	1110	2 I51116	NF-180 - sea lamp
4	188.5	21.2	1020	1 QPHUH	neurofilament trip
5	186	20.9	335	2 T33457	hypothetical prote
6	183	20.6	1701	2 T09127	probable erythrocy
7	182.5	20.5	771	1 A33430	h-caldesmon - chic
8	181.5	20.4	312	2 T25994	hypothetical prote
9	181.5	20.4	1403	2 T11583	probable translati
10	180	20.2	606	2 A43427	neurofilament trip
11	178.5	20.1	332	2 B43427	neurofilament trip
12	178.5	20.1	1038	2 JCS487	claustrin - chicke
13	177	19.9	1087	1 QPM5H	neurofilament trip
14	176.5	19.9	644	2 S55395	neurofilament trip
15	176.5	19.9	1072	1 A37221	neurofilament trip
16	176	19.8	762	2 G88436	protein T04A8.13
17	176	19.8	791	2 T24435	hypothetical prote
18	176	19.8	1178	2 S78475	mannosylphosphoryl
19	175	19.7	1052	1 A44937	kinetoplast-associ
20	173	19.5	1390	2 S51384	sperm tail-specific
21	172.5	19.4	854	2 S02003	neurofilament trip
22	172	19.3	1132	2 T43483	translational initia
23	171.5	19.3	805	2 E70474	translational initia
24	170.5	19.2	344	2 S34153	met101-1 protein -
25	170.5	19.2	532	2 C96608	hypothetical prote
26	170.5	19.2	798	2 I50479	neurofilament medi
27	169.5	19.1	385	2 T19201	hypothetical prote
28	169.5	19.1	699	2 E84565	hypothetical prote
29	169	19.0	1359	2 T34036	hypothetical prote

30	168.5	19.0	407	1 EDBEQ3	immediate-early pr
31	168.5	19.0	1192	2 A71623	probable secreted
32	167	18.8	5327	2 T13564	microtubule-associ
33	166.5	18.7	845	2 A45659	neurofilament trip
34	166.5	18.7	6642	2 T29757	neurofilament trip
35	164	18.4	1526	2 A45605	protein UNC-89 - C
36	163.5	18.4	2364	2 A56577	mature-parasite-in
37	162.5	18.3	784	2 PN0009	microtubule-associ
38	162	18.2	491	2 C97267	neurofilament trip
39	160	18.0	667	2 A40713	hypothetical prote
40	159.5	17.9	210	2 T28771	cylicin I - bovine
41	159.5	17.9	721	2 A33319	hypothetical prote
42	159	17.9	849	2 S00030	microtubule-associ
43	158.5	17.8	619	2 A41971	neurofilament trip
44	158.5	17.8	619	2 A97887	surface protein ps
45	158.5	17.8	806	2 T23648	hypothetical prote

## ALIGNMENTS

```
RESULT 1
H95255
choiline binding protein A [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: H95255
R/tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfe,
nson, T.; Hickey, E.K.; Holt, I.E.
A/Authors: loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A./Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; M01D:21357209; PMID:11463916
A/Accession: H95255
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-693 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK76241.1; PID:g14973701; GSPDB:GN00164; TIGR:SPA
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP2190

Query Match      100.0%; Score 889; DB 2; Length 693;
Best Local Similarity 100.0%; Pred. No. 2.6e-38;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IKTDREBEAEAKRRADAKEOGKPKGRAGVGEIATPPKENDAKSSDSVGEETLPS 60
        |||||||
Db      265 IKTDREBEAEAKRRADAKEOGKPKGRAGVGEIATPPKENDAKSSDSVGEETLPS 324

QY      61 PSIKPEKVAEAKVEAKKKAEDKEDRRNYPNTYKTLLEIAESDVEYKAELEL 120
        |||||||
Db      325 PSIKPEKVAEAKVEAKKKAEDKEDRRNYPNTYKTLLEIAESDVEYKAELEL 364

QY      121 VKEAKEPRNEEVYKAKAEVSKKAEATRLKIKTKDKKAEBAKKAEDKVEKPA 180
        |||||||
Db      385 VKEAKEPRNEEVYKAKAEVSKKAEATRLKIKTKDKKAEBAKKAEDKVEKPA 444

RESULT 2
H98120
choiline binding protein A [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: H98120
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bugeff, S.; Dehoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.O.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.W.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A./Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; M01D:21429245; PMID:11544234
```

A:Accession H98120  
A:Status: preliminary  
Molecule type: DNA  
A:Residues: 1701  
A:Cross-references: GB:AE007317, FIDN:AAI00797.1, PID:g15459699, GSEDB:GN00174  
C:Genetics: pspc  
A:Gene: pspc

Query Match	90.4%;	Score 803.5;	DB 2;	Length 701;
Best Local Similarity	86.8%;	Pred. No. 5.4e-34;		
Matches 164;	Conservative 12;	Mismatches 4;	Indels 9;	Gaps 1;

```

0y      1 IKTDEEAEAEKRRADK-----EGKGKGKAGVPEPLATPDKKENDAKSDS 51
Db      226 IKTDRKAEAEKRRADKLTKEANVATSDGKGGKRGVPEPLATPDKKENDAKSDS 285
0y      52 SVGEETLPSPSLPEKKVAEAEKVEAEKKAADQKEEDRRNPPTNTYKTLAEIASDV 111
Db      266 SVGEETLPSSLSKGGKVAEAEKVEAEKKAADQKEEDRRNPPTNTYKTLDEIASDV 345
0y      112 EYKKAELLEVKEAEKPRNEEKVQAKAEVSKAEATRLRKIKTDKKAEEAKRFAAE 171
Db      346 KYKKAELLEVKEAEKPRDEBKIKQAKAVESKKAETRLRNIKTDRKKAEEAKRFAAE 405

```

[illegible]

RESULT 3  
I51116  
NF-180 - sea lamprey  
C:Species: Petromyzon marinus (sea lamprey)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: I51116  
R:Jacobs, A.J.; Kamholz, J.; Seizer, M.E.  
Brain Res. Mol. Brain Res. 29, 43-52, 1995  
A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re  
A:Reference number: I51116; MUID:95287814; PMID:7770000  
A:Accession: I51116  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1110 <JUNC>  
A:Cross-references: EMBL:U19361; NID:G632548; PIDD:AAA0106.1; PID:G632549  
A:Superfamily: neurofilament triplet H protein

Query Match	21.8%;	Score 194;	DB 2;	Length 1110;
-------------	--------	------------	-------	--------------

Best Local Similarity 29.6%; Pred. No. 0.0052;  
Matches 58; Conservative 38; Mismatches 74; Indels 26; Gaps 5;

```

QY      2  TTTDREAEAEARRRDAKQCKPRKGAKGVPELMTTPDCKENDAKSSDSVGEETLPSF  61
Db      527  EKEKEKEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE  580
QY      62  SLKPEKVAEAKKVEAEAKKKAEDQKEDBRNRPNTYKTLLEIAESDVEAKAELTV  121
Db      561  TEKAAAEAKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE  639
QY      122  KEEAEPR-NEEKVQAKAEVSKKAETRL-----EKIKTRKKAEEBAKRK  168
Db      640  EEEAEEEVTSKAKTQAEAEVEEEAEAAAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE  699
QY      169  A-----AEEDKVKEK  178
Db      700  ADAEEDAEAEVEEVKEE  715

```

RESULT 4  
QPRUH  
neurofilament triplet H protein - human  
N:Alternate names: neurofilament protein, 112K  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jun-2000

C:Accession: S00979  
R:Rees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazarini, R.A.  
EMBO J. 7 1947-1955, 1988  
A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)  
A:Reference number: S00979, MUID:86528581; PMID:3138108  
A:Accession: S00979  
A:Molecule type: DNA

A/Cross-references: EMBL:X15306, NID:g35028, PIDN:CAA33366.1, PID:g1841430  
A/Note: it is uncertain whether Met-1 or Met-2 is the initiator  
C/Genetics:

A:Cross-references: GDB:120225; OMIM:162230  
A:Map position: 22q12.1-22q13.1  
A:Introns: 295/1, 361/3, 403/2  
C:Superfamily: neurofilament triplet H protein  
C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein  
F:1-100/Domain: amino-terminal <NTD>  
F:101-410/Domain: rod #status predicted <ROD>  
F:411-1020/Domain: carboxyl-terminal <CTD>  
F:502-826/Region: 14-residue repeats  
F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,646,652,658,664,670,676,682,688,694,700,706,712,718,724,730,736,742,748,754,760,766,772,778,784,790,796,802,808,814,820,826,832,838,844,850,856,862,868,874,880,886,892,898,904,910,916,922,928,934,940,946,952,958,964,970,976,982,988,994,1000,1006,1012,1018,1024,1030,1036,1042,1048,1054,1060,1066,1072,1078,1084,1090,1096,1102,1108,1114,1120,1126,1132,1138,1144,1150,1156,1162,1168,1174,1180,1186,1192,1198,1204,1210,1216,1222,1228,1234,1240,1246,1252,1258,1264,1270,1276,1282,1288,1294,1300,1306,1312,1318,1324,1330,1336,1342,1348,1354,1360,1366,1372,1378,1384,1390,1396,1402,1408,1414,1420,1426,1432,1438,1444,1450,1456,1462,1468,1474,1480,1486,1492,1498,1504,1510,1516,1522,1528,1534,1540,1546,1552,1558,1564,1570,1576,1582,1588,1594,1600,1606,1612,1618,1624,1630,1636,1642,1648,1654,1660,1666,1672,1678,1684,1690,1696,1702,1708,1714,1720,1726,1732,1738,1744,1750,1756,1762,1768,1774,1780,1786,1792,1798,1804,1810,1816,1822,1828,1834,1840,1846,1852,1858,1864,1870,1876,1882,1888,1894,1900,1906,1912,1918,1924,1930,1936,1942,1948,1954,1960,1966,1972,1978,1984,1990,1996,2002,2008,2014,2020,2026,2032,2038,2044,2050,2056,2062,2068,2074,2080,2086,2092,2098,2104,2110,2116,2122,2128,2134,2140,2146,2152,2158,2164,2170,2176,2182,2188,2194,2200,2206,2212,2218,2224,2230,2236,2242,2248,2254,2260,2266,2272,2278,2284,2290,2296,2302,2308,2314,2320,2326,2332,2338,2344,2350,2356,2362,2368,2374,2380,2386,2392,2398,2404,2410,2416,2422,2428,2434,2440,2446,2452,2458,2464,2470,2476,2482,2488,2494,2500,2506,2512,2518,2524,2530,2536,2542,2548,2554,2560,2566,2572,2578,2584,2590,2596,2602,2608,2614,2620,2626,2632,2638,2644,2650,2656,2662,2668,2674,2680,2686,2692,2698,2704,2710,2716,2722,2728,2734,2740,2746,2752,2758,2764,2770,2776,2782,2788,2794,2800,2806,2812,2818,2824,2830,2836,2842,2848,2854,2860,2866,2872,2878,2884,2890,2896,2902,2908,2914,2920,2926,2932,2938,2944,2950,2956,2962,2968,2974,2980,2986,2992,2998,3004,3010,3016,3022,3028,3034,3040,3046,3052,3058,3064,3070,3076,3082,3088,3094,3100,3106,3112,3118,3124,3130,3136,3142,3148,3154,3160,3166,3172,3178,3184,3190,3196,3202,3208,3214,3220,3226,3232,3238,3244,3250,3256,3262,3268,3274,3280,3286,3292,3298,3304,3310,3316,3322,3328,3334,3340,3346,3352,3358,3364,3370,3376,3382,3388,3394,3400,3406,3412,3418,3424,3430,3436,3442,3448,3454,3460,3466,3472,3478,3484,3490,3496,3502,3508,3514,3520,3526,3532,3538,3544,3550,3556,3562,3568,3574,3580,3586,3592,3598,3604,3610,3616,3622,3628,3634,3640,3646,3652,3658,3664,3670,3676,3682,3688,3694,3700,3706,3712,3718,3724,3730,3736,3742,3748,3754,3760,3766,3772,3778,3784,3790,3796,3802,3808,3814,3820,3826,3832,3838,3844,3850,3856,3862,3868,3874,3880,3886,3892,3898,3904,3910,3916,3922,3928,3934,3940,3946,3952,3958,3964,3970,3976,3982,3988,3994,4000,4006,4012,4018,4024,4030,4036,4042,4048,4054,4060,4066,4072,4078,4084,4090,4096,4102,4108,4114,4120,4126,4132,4138,4144,4150,4156,4162,4168,4174,4180,4186,4192,4198,4204,4210,4216,4222,4228,4234,4240,4246,4252,4258,4264,4270,4276,4282,4288,4294,4300,4306,4312,4318,4324,4330,4336,4342,4348,4354,4360,4366,4372,4378,4384,4390,4396,4402,4408,4414,4420,4426,4432,4438,4444,4450,4456,4462,4468,4474,4480,4486,4492,4498,4504,4510,4516,4522,4528,4534,4540,4546,4552,4558,4564,4570,4576,4582,4588,4594,4600,4606,4612,4618,4624,4630,4636,4642,4648,4654,4660,4666,4672,4678,4684,4690,4696,4702,4708,4714,4720,4726,4732,4738,4744,4750,4756,4762,4768,4774,4780,4786,4792,4798,4804,4810,4816,4822,4828,4834,4840,4846,4852,4858,4864,4870,4876,4882,4888,4894,4900,4906,4912,4918,4924,4930,4936,4942,4948,4954,4960,4966,4972,4978,4984,4990,4996,5002,5008,5014,5020,5026,5032,5038,5044,5050,5056,5062,5068,5074,5080,5086,5092,5098,5104,5110,5116,5122,5128,5134,5140,5146,5152,5158,5164,5170,5176,5182,5188,5194,5200,5206,5212,5218,5224,5230,5236,5242,5248,5254,5260,5266,5272,527

Query Match	21.2%;	Score 188.5;	DB 1;	Length 1020;
Best Local Similarity	30.9%;	Pred. No. 0.0092;		
Matches 58; Conservative	28;	Mismatches 79;	Indels 23;	Gaps 6;

```

QY 5 REAEAEKKRRADAEOKPKGRARAGVDELATDCKKENDKSSSSSGEBTLPSPSLK 64
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 KEAEASPEKASPVVEAEKSPPEKASPVVEAEKTEBKASPVVEAEKS-----PEKAS 752
QY 65 PEKTVAEAEKVEAEKKAEDOKEDRRNYPINTYKITLELELAESD-----VEKXAE 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 753 PE-KKTLTIVSSPEKTPPEKAEARSADDFPEKASPVVEAEKSPPEKASPLKADKAAPE 811
QY 118 LEL-VKEAEKEBRNE-----KYQAKAEVSKSAEATRLTEKIKTDKRAABEAKRAA 170
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 812 KEIPKKEEVKSPVKEEERQVEKVEPEPKKAEAEKAPAPKTEKKDSK--BEAEKKEA 869

```

QY	171	EEDKYREK	178
	:	:	
Db	870	PKPKYEEK	877

RESULT 5  
 T33457  
 hypochromic protein F36H12.3 - *Caenorhabditis elegans*  
 C1Species: *Caenorhabditis elegans*  
 C1Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #ext\_change 29-Oct-1999  
 C1Accession: T33457  
 R1Blanchard, M.; Bradshaw, H.; Stellyes, L.  
 submitted to the EMBL Data Library, July 1998  
 A1Description: The sequence of C. elegans cosmid F36H12.  
 A1Reference number: Z21346  
 A1Accession: T33457  
 A1Status: preliminary; translated from GB/EMBL/DBJ  
 A1Molecule type: DNA  
 A1Residues: 1-335 <BLA>  
 A1Cross-references: EMBL:AF078790; P1DN:AC26930.1; GSPDB:GN00022; CESP:F36H12.3  
 A1Experimental source: strain Bristol N2; clone F36H12  
 C1Genetics:  
 A1Gene: CESP:F36H12.3  
 A1Map position: 4  
 A1Introns: 32/3; 227/1; 270/2

```
Query Match      20.9% Score 186; DB 2; Length 335;
Best Local Similarity 31.1% Pred. No. 0.0047;
Matches 56; Conservative 31; Mismatches 77; Indels 16; Gaps 5;
```











```

QY 1 KTBREAEAEKKRPAADAEQGGKPKRAKRGVGPLAMPD-KKENDADSSSSVGEETLP 59
Db 268 VSVKEEKEEBEABGKEEBQEAEEVAANKKS-PVATTTPEIKEEBGEKEEQQEBEEEEE 326
QY 60 SPSLK-----PEKVAEAEKKVEEA-----KKKAEDQKEE-DRNRY 94
Db 327 DEGVASDQAEBGEGSEKSGSKNKGEGEGEFTAEABGEVEDEAEAKKEKTEKSEVAAKEE 386
QY 95 PNTYKTELELEAEEDVEVKKAELELVKEEAKEPNEEKVKQAKAEVSSKA--EATLE 152
Db 387 PVTEAKVGKPEAKAKSPVPSQVPEEYVAPRAEATAGGDEQKEEBEKEEKKAAKESPKEE 446
QY KIK-----TDRKAAEEBAKRAAEADKVYKEP 179
Db 447 KVEKKEEKRPDVPKKGAESPVEKEEAEEAATITTR 481

```

## RESULT 15

neurofilament triplet H protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Dec-1999  
C:Accession: A37221; A25649; A32757; B25649  
C:Chn, S.S.M.; Liem, R.K.H.  
J. Neurosci. 10, 3714-3726, 1990  
A:Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vimentin  
A:Reference number: A37221; PMID:91038277; PMID:2230956  
A:Accession: A37221  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1072 <CHI>  
A:Cross-references: GB:AF031879; NID:g2642597; PIDN:AA87068.1.; PID:g2642598  
R:Robinson, P.A.; Wiom, D.; Anderson, B.H.  
FEBS Lett. 209, 203-205, 1986  
A:Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).  
A:Reference number: A25649; PMID:87080760; PMID:2878828  
A:Accession: A25649  
A:Molecule type: mRNA  
A:Residues: 230-318;472-542 <ROB>  
A:Cross-references: GB:M37227  
R:Dautigny, A.; Plam-Dinn, D.; Roussel, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.  
Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988  
A:Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection  
A:Reference number: A30796; PMID:88309090; PMID:2457365  
A:Accession: A30796  
A:Molecule type: mRNA  
A:Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-777, 'S', 779-827, 'S', 829-830, 'S', 831-832, 'S', 833-834, 'S', 835-836, 'S', 837-838, 'S', 839-840, 'S', 841-842, 'S', 843-844, 'S', 845-846, 'S', 847-848, 'S', 849-850, 'S', 851-852, 'S', 853-854, 'S', 855-856, 'S', 857-858, 'S', 859-860, 'S', 861-862, 'S', 863-864, 'S', 865-866, 'S', 867-868, 'S', 869-870, 'S', 871-872, 'S', 873-874, 'S', 875-876, 'S', 877-878, 'S', 879-880, 'S', 881-882, 'S', 883-884, 'S', 885-886, 'S', 887-888, 'S', 889-890, 'S', 891-892, 'S', 893-894, 'S', 895-896, 'S', 897-898, 'S', 899-900, 'S', 901-902, 'S', 903-904, 'S', 905-906, 'S', 907-908, 'S', 909-910, 'S', 911-912, 'S', 913-914, 'S', 915-916, 'S', 917-918, 'S', 919-920, 'S', 921-922, 'S', 923-924, 'S', 925-926, 'S', 927-928, 'S', 929-930, 'S', 931-932, 'S', 933-934, 'S', 935-936, 'S', 937-938, 'S', 939-940, 'S', 941-942, 'S', 943-944, 'S', 945-946, 'S', 947-948, 'S', 949-950, 'S', 951-952, 'S', 953-954, 'S', 955-956, 'S', 957-958, 'S', 959-960, 'S', 961-962, 'S', 963-964, 'S', 965-966, 'S', 967-968, 'S', 969-970, 'S', 971-972, 'S', 973-974, 'S', 975-976, 'S', 977-978, 'S', 979-980, 'S', 981-982, 'S', 983-984, 'S', 985-986, 'S', 987-988, 'S', 989-990, 'S', 991-992, 'S', 993-994, 'S', 995-996, 'S', 997-998, 'S', 999-1000, 'S', 1001-1002, 'S', 1003-1004, 'S', 1005-1006, 'S', 1007-1008, 'S', 1009-1010, 'S', 1011-1012, 'S', 1013-1014, 'S', 1015-1016, 'S', 1017-1018, 'S', 1019-1020, 'S', 1021-1022, 'S', 1023-1024, 'S', 1025-1026, 'S', 1027-1028, 'S', 1029-1030, 'S', 1031-1032, 'S', 1033-1034, 'S', 1035-1036, 'S', 1037-1038, 'S', 1039-1040, 'S', 1041-1042, 'S', 1043-1044, 'S', 1045-1046, 'S', 1047-1048, 'S', 1049-1050, 'S', 1051-1052, 'S', 1053-1054, 'S', 1055-1056, 'S', 1057-1058, 'S', 1059-1060, 'S', 1061-1062, 'S', 1063-1064, 'S', 1065-1066, 'S', 1067-1068, 'S', 1069-1070, 'S', 1071-1072, 'S', 1073-1074, 'S', 1075-1076, 'S', 1077-1078, 'S', 1079-1080, 'S', 1081-1082, 'S', 1083-1084, 'S', 1085-1086, 'S', 1087-1088, 'S', 1089-1090, 'S', 1091-1092, 'S', 1093-1094, 'S', 1095-1096, 'S', 1097-1098, 'S', 1099-1100, 'S', 1101-1102, 'S', 1103-1104, 'S', 1105-1106, 'S', 1107-1108, 'S', 1109-1110, 'S', 1111-1112, 'S', 1113-1114, 'S', 1115-1116, 'S', 1117-1118, 'S', 1119-1120, 'S', 1121-1122, 'S', 1123-1124, 'S', 1125-1126, 'S', 1127-1128, 'S', 1129-1130, 'S', 1131-1132, 'S', 1133-1134, 'S', 1135-1136, 'S', 1137-1138, 'S', 1139-1140, 'S', 1141-1142, 'S', 1143-1144, 'S', 1145-1146, 'S', 1147-1148, 'S', 1149-1150, 'S', 1151-1152, 'S', 1153-1154, 'S', 1155-1156, 'S', 1157-1158, 'S', 1159-1160, 'S', 1161-1162, 'S', 1163-1164, 'S', 1165-1166, 'S', 1167-1168, 'S', 1169-1170, 'S', 1171-1172, 'S', 1173-1174, 'S', 1175-1176, 'S', 1177-1178, 'S', 1179-1180, 'S', 1181-1182, 'S', 1183-1184, 'S', 1185-1186, 'S', 1187-1188, 'S', 1189-1190, 'S', 1191-1192, 'S', 1193-1194, 'S', 1195-1196, 'S', 1197-1198, 'S', 1199-1200, 'S', 1201-1202, 'S', 1203-1204, 'S', 1205-1206, 'S', 1207-1208, 'S', 1209-1210, 'S', 1211-1212, 'S', 1213-1214, 'S', 1215-1216, 'S', 1217-1218, 'S', 1219-1220, 'S', 1221-1222, 'S', 1223-1224, 'S', 1225-1226, 'S', 1227-1228, 'S', 1229-1230, 'S', 1231-1232, 'S', 1233-1234, 'S', 1235-1236, 'S', 1237-1238, 'S', 1239-1240, 'S', 1241-1242, 'S', 1243-1244, 'S', 1245-1246, 'S', 1247-1248, 'S', 1249-1250, 'S', 1251-1252, 'S', 1253-1254, 'S', 1255-1256, 'S', 1257-1258, 'S', 1259-1260, 'S', 1261-1262, 'S', 1263-1264, 'S', 1265-1266, 'S', 1267-1268, 'S', 1269-1270, 'S', 1271-1272, 'S', 1273-1274, 'S', 1275-1276, 'S', 1277-1278, 'S', 1279-1280, 'S', 1281-1282, 'S', 1283-1284, 'S', 1285-1286, 'S', 1287-1288, 'S', 1289-1290, 'S', 1291-1292, 'S', 1293-1294, 'S', 1295-1296, 'S', 1297-1298, 'S', 1299-1300, 'S', 1301-1302, 'S', 1303-1304, 'S', 1305-1306, 'S', 1307-1308, 'S', 1309-1310, 'S', 1311-1312, 'S', 1313-1314, 'S', 1315-1316, 'S', 1317-1318, 'S', 1319-1320, 'S', 1321-1322, 'S', 1323-1324, 'S', 1325-1326, 'S', 1327-1328, 'S', 1329-1330, 'S', 1331-1

Query Match	19.9%	Score 176.5;	DB 1;	Length 1072;
Best Local Similarity	28.0%	Pred. No. 0.039;		
Matches 59; Conservative	28;	Mismatches 71;	Indels 53;	Gaps 6

QY 1 EARRPDAKGGQPKRKARGVGELATPDKKENAKSDSVGEETLPSPSLPEKKVA 70

Db 735 EAKSPAEAKSPAEAKSPAEKSPVVEKSPKAKSPVKGCAKSLAEKSPKAKSPVVEEI 794

QY 71 EAERKVEAKKAAEDKEDRRNPNTYKTELEIAESDVEVKAELELYKEAKRP-- 128

Db 795 KPAEAYKSPKAKSPKKEAK--SPKAKTLDVSPKAKRP-----AKKAKPAD 843

```

QY      129  -RNEEKYQ-AKAEVSKAAEATREK-----KTGRKAAEE-----164
          . . . . . : . . . . .
Db      844  IREPEQVSKAKAEASPEKETREKVAAPKKEEVSPVEEVAKKEPKVVEEKTPATP  903
          . . . . . : . . . . .
QY      165  -----AKRAAEEDKVEKP  179
          . . . . . : . . . . .
Db      904  KTEVSKSKDEAPKAQPKPAEKKEPPLTEK  934
          . . . . . : . . . . .

```

Search completed: November 21, 2003, 13:34:30  
Job time : 8.29378 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:28:52 ; Search time 16.6715 Seconds  
(without alignments)  
2786.162 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442  
Perfect score: 889  
Sequence: 1 IKTDREAEAEAKRADAKE.....ABEEAKRAEDDKVEKPA 180

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	516	2 Q9ROT2	Q9ROT2 streptococc
2	889	100.0	693	16 Q97A74	Q97A74 streptococc
3	880	99.0	693	2 Q9KK21	Q9KK21 streptococc
4	871	98.0	696	2 Q9KK32	Q9KK32 streptococc
5	866.5	97.5	752	2 Q9KK30	Q9KK30 streptococc
6	816.5	91.8	655	2 Q9KK50	Q9KK50 streptococc
7	815.5	91.7	680	2 Q9KK43	Q9KK43 streptococc
8	814.5	91.6	680	2 Q9KK11	Q9KK11 streptococc
9	809	91.0	660	2 Q9KK18	Q9KK18 streptococc
10	807	90.8	730	2 Q9KK47	Q9KK47 streptococc
11	804	90.4	695	2 Q9KK36	Q9KK36 streptococc
12	803.5	90.4	564	2 Q69188	Q69188 streptococc
13	803.5	90.4	581	2 Q33742	Q33742 streptococc
14	803.5	90.4	663	2 Q30874	Q30874 streptococc
15	803.5	90.4	701	2 Q9KK48	Q9KK48 streptococc
16	803.5	90.4	701	16 Q8DN05	Q8DN05 streptococc

17	803	90.3	681	2 Q9KK16	Q9KK16 streptococc
18	799.5	89.9	701	2 Q9ROT5	Q9ROT5 streptococc
19	795	89.4	730	2 Q9KK31	Q9KK31 streptococc
20	794.5	89.4	657	2 Q9KK29	Q9KK29 streptococc
21	793.5	89.3	709	2 Q9KK38	Q9KK38 streptococc
22	787	88.5	681	2 Q9KK13	Q9KK13 streptococc
23	779	87.6	678	2 Q9KK49	Q9KK49 streptococc
24	776	87.3	650	2 Q9KK41	Q9KK41 streptococc
25	772.5	86.9	695	2 Q9KK51	Q9KK51 streptococc
26	770.5	86.7	659	2 Q9KK42	Q9KK42 streptococc
27	761	85.6	684	2 Q9KK17	Q9KK17 streptococc
28	757.5	85.2	684	2 Q9KK46	Q9KK46 streptococc
29	752	84.6	487	2 Q9ROT3	Q9ROT3 streptococc
30	751.5	84.5	752	2 Q9KK33	Q9KK33 streptococc
31	728.5	81.9	488	2 Q9ROT6	Q9ROT6 streptococc
32	716.5	80.6	676	2 Q9KK39	Q9KK39 streptococc
33	716.5	80.6	696	2 Q9KK14	Q9KK14 streptococc
34	709.5	79.8	739	2 Q9ROT4	Q9ROT4 streptococc
35	704.5	79.2	869	2 Q9KK27	Q9KK27 streptococc
36	702.5	79.0	678	2 Q9KK54	Q9KK54 streptococc
37	693.5	78.0	565	2 Q9KK15	Q9KK15 streptococc
38	684.5	77.0	820	2 Q9ROT1	Q9ROT1 streptococc
39	684.5	77.0	929	2 Q9ZAV5	Q9ZAV5 streptococc
40	684.5	77.0	929	2 Q9KK19	Q9KK19 streptococc
41	551.5	62.0	523	2 Q33753	Q33753 streptococc
42	521	58.6	107	2 Q93MU1	Q93MU1 streptococc
43	517	58.2	107	2 Q93MU4	Q93MU4 streptococc
44	516.5	58.1	667	2 Q9KK28	Q9KK28 streptococc
45	514	57.8	539	2 Q33741	Q33741 streptococc

# ALIGNMENTS

## RESULT 1

ID Q9ROT2 PRELIMINARY; PRT; 516 AA.

AC Q9ROT2; 01-MAY-2000 (TREMREL. 13, Created)

DT 01-MAY-2000 (TREMREL. 13, Last sequence update)

DE 01-MAR-2003 (TREMREL. 23, Last annotation update)

HT Hypothetical 57.9 kDa protein (Fragment).

GN PSpC.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=181905;

RX MEDLINE=20038319; PubMed=10569772;

RA Brooks-Walter A., Biles D.E., Hollingshead S.K.;

RT "The pspC gene of streptococcus pneumoniae encodes a polymorphic

RT protein, PSpC, which elicits cross-reactive antibodies to PspA and

RT provides immunity to pneumococcal bacteremia.";

RL Infect. Immun. 67:6533-6542(1999).

DR EMBL; AF068649; AAF13459.1; --

DR Interpro; IPR005877; Gpos\_YSTRK.

DR Pfam; PFO4650; YSTRK\_signal; 1.

KW TIGRPFAMs; TIGR01168; YSTRK\_signal; 1.

FT Hypothetical protein.

FT NON TER 516

SO SEQUENCE 516 AA; 57886 MW; 1119782688B7E35 CRC64;

Query Match 100.0%; Score 889; DB 2; Length 516;

Best Local Similarity 100.0%; Pred. No. 2.4e-42; Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKEQKPKGRKRGVGEGLATPPKKNDAKSSNSVGEETLPS 60  
DB 264 IKTDREAEAEAKRADAKEQKPKGRKRGVGEGLATPPKKNDAKSSNSVGEETLPS 323  
QY 61 PSLKPKKVAEAKKVEAKKAEDQKEDRRNYPNTYKTELEIAESDVEVKAELEL 120

Db 324 PSLEKKEVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 383  
 Oy 121 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 180  
 Db 384 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 443

## RESULT 2

O97N74 PRELIMINARY; PRT; 693 AA.  
 ID O97N74  
 AC O97N74;  
 DT 01-OCT-2001 (Tremblrel. 18, Created)  
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Choline binding protein A.  
 GN SP2190.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21357209; PubMed=11463916;  
 RX STRAIN=TIGR4;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 Peterson S., Heideberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 Durkin A.S., Gwin W., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,  
 Holtzaple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,  
 McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,  
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";  
 RL Science 293:498-506(2001).  
 DR EMBL; AE007507; AAK76241.1; -  
 DR TIGR; SP2190; -  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003877; Gpos\_Ysirk.  
 DR Pfam; PF01473; CW\_binding\_1; 8.  
 DR Pfam; PF04650; Ysirk\_signal; 1.  
 DR TIGRFAMs; TIGR01168; Ysirk\_signal; 1.  
 KW Complete proteome.

SO SEQUENCE 693 AA; 77762 MW; 6F8F47C32E344A41 CRC64;

Query Match 100.0%; Score 889; DB 16; Length 693;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-42;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
 Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 324  
 Oy 61 PSLEKKEVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 120  
 Db 325 PSLEKKEVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 384  
 Oy 121 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 180  
 Db 385 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 444

## RESULT 3

O9KK21 PRELIMINARY; PRT; 693 AA.  
 ID O9KK21  
 AC O9KK21;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Surface protein PspC.  
 GN PspC.  
 OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=g9;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
 Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154035; AAF73807.1; -  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR005877; Gpos\_Ysirk.  
 DR Pfam; PF01473; CW\_binding\_1; 8.  
 DR Pfam; PF04650; Ysirk\_signal; 1.  
 DR TIGRFAMs; TIGR01168; Ysirk\_signal; 1.  
 SO SEQUENCE 693 AA; 77789 MW; 8DA1BA115665667 CRC64;

Query Match 99.0%; Score 880; DB 2; Length 693;  
 Best Local Similarity 99.4%; Pred. No. 1e-41;  
 Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
 Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 324  
 Oy 61 PSLEKKEVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 120  
 Db 325 PSLEKKEVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 384  
 Oy 121 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 180  
 Db 385 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKKAEEAKRAAEEDKVKEXPA 444

## RESULT 4

O9KK32 PRELIMINARY; PRT; 696 AA.  
 ID O9KK32  
 AC O9KK32;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Surface protein PspC.  
 GN PspC.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=g402;  
 RA Iannelli F., Oggioni M.R., Pozzi G.;  
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
 Streptococcus pneumoniae";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF154027; AAF73796.1; -  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR005877; Gpos\_Ysirk.  
 DR Pfam; PF01473; CW\_binding\_1; 8.  
 DR Pfam; PF04650; Ysirk\_signal; 1.  
 DR TIGRFAMs; TIGR01168; Ysirk\_signal; 1.  
 SO SEQUENCE 696 AA; 78444 MW; DFA15A16166D7D9 CRC64;

Query Match 98.0%; Score 871; DB 2; Length 696;  
 Best Local Similarity 97.3%; Pred. No. 3.2e-41;  
 Matches 179; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

Oy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60  
 Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 324  
 Oy 57 TLSPSLKPEKVAEAEKVEAEKKAEDQKEDRRNRYTNTYKLELEIASDVEVKAELEL 116

Db 325 TLPSPLKEPKVAEAEKKEAEKKAADQKEEDRRNYPNTYKTLLELEIASDVEVKKA 384  
Qy 117 ELELVKEAEKPEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAEEDVK 176  
Db 385 ELELVKEAEKPEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAEEDVK 444  
Qy 177 EKPA 180  
Db 445 EKPA 448

## RESULT 5

Q9KK30 PRELIMINARY; PRT; 752 AA.

AC Q9KK30  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9408;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154029; AAF73798.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW\_binding\_1; 12.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 752 AA; 85296 MW; D3E50B8CFCABD3 CRC64;

Query Match 97.5%; Score 866.5; DB 2; Length 752;  
Best Local Similarity 94.2%; Pred. No. 6.1e-41;  
Matches 178; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEKRRADAK-----QCKPKRAKRGVPELATPPDKKENDAKSSDS 51  
Db 237 IKTDREAEAEKRRADAKKEAEADSDQCKPKRAKRGVPELATPPDKKENDAKSSDS 296  
Qy 52 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDV 111  
Db 297 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDV 356  
Qy 112 EVKAELELVKEAEKPEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAE 171  
Db 357 EVKAELELVKEAEKPEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAE 416  
Qy 172 EDKVEKPA 180  
Db 417 EDKVEKPA 425

## RESULT 6

Q9KK50 PRELIMINARY; PRT; 655 AA.

AC Q9KK50  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=8R1;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154010; AAF73777.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR InterPro; IPR00533; Tropomyosin.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR PRINTS; PR00194; TROPOMYOSIN.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 655 AA; 73951 MW; 30623F1EDB9D05C2 CRC64;

Query Match 91.8%; Score 816.5; DB 2; Length 655;  
Best Local Similarity 93.3%; Pred. No. 3.3e-38;  
Matches 168; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Qy 1 IKTDREAEAEKRRADAKQCKPKRAKRGVPELATPPDKKENDAKSSDSVGEETLPS 60  
Db 228 IKTERKKAEEBAKRRADAKQCKPKRAKRGVPEQATPPDKKENDAKSSDSVGEETLPN 287  
Qy 61 PSLLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDVEVKKAEL 120  
Db 288 PSLLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDVEVKKAEL 347  
Qy 121 VKEAEKPEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAEEDVKKPA 180  
Db 348 VKEAEKSRNEKKEVKQAQAEVSKAEATRLKIKTRKKAEEBAKRAAEEDVKKPA 406

## RESULT 7

Q9KK43 PRELIMINARY; PRT; 680 AA.

AC Q9KK43  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9376;  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154016; AAF73784.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW\_binding\_1; 8.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 680 AA; 76727 MW; AAD51C2A31495B8E CRC64;

Query Match 91.7%; Score 815.5; DB 2; Length 680;  
Best Local Similarity 88.4%; Pred. No. 3.8e-38;  
Matches 167; Conservative 10; Mismatches 3; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEKRRADAK-----EGCKPKRAKRGVPELATPPDKKENDAKSSDS 51  
Db 237 IKTDREAEAEKRRADAKKEANVATSDGCKPKRAKRGVPELATPPDKKENDAKSSDS 296  
Qy 52 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDV 111  
Db 297 SVGEETLPSSSLSGKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLELEIASDV 356

OY	112 EVKKAELVELVKEAKPEPNDEKVKOKAAEVESKKAEATRLKXIKTDRCKAAEEBKRAAE	172
Dd	357 KVKKABELEVKEBAKPPOBSEKI KQAKAVESKQAEATRLEKIKTDRCCKAAEBEKRRDAE	416
OY	172 EDKVKEKPA	180
Dd	417 EDKVKKEPA	425

## RESULT 8

ID	Q9KK11	PRELIMINARY;	PRT;	680 AA
----	--------	--------------	------	--------

DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxId=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Iannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus *pspC* of  
RT Streptococcus pneumoniae.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF154045; AAF73822.1; -  
DR InterPro; IPR002479; CM binding  
DR InterPro; IPR005877; Gpos YSIRK.  
DR Pfam; PF01473; CM binding\_1; 9.  
DR Pfam; PF04650; YSIRK signal; 1.  
DR TIGRFAMs; TIGR01166; YSIRK signal; 1.  
SQ SEQUENCE 680 AA; 76185 MW; 07D682193DD0A0F9 CRC64;

Query Match 91.6%; Score 814.5; DB 2; Length 680;

Matches 171; Conservative 3; Mismatches 4; Indels 9; Gaps 2,

Qy	1	IKTDRREAEAEARRRDAE-----OGFKGAKKGVGELATPDKENDAKSSDS	52
Db	222	IKTDRREAEAEARRRDAEQEANVATSGODESKRAKGVGELATPDKENDAKSSDS	261
Qy	53	VGEETLPSPSLKPEKTVAAEKKEVEBAKKAADQKEEDRRNPNTYKTLLEIAESDVE	112
Db	282	VGEETLPSPSLKPEKTVAAEKKEVEBAKKAADQKEEDRRNPNTYKTLLEIAESDVE	341
Qy	113	VKAQELVYKEAEKPRNEEKYKQAKAEVSKKAEATLEKIKTDRKAAEBAARKAAE	172
Db	342	VKAQELVYKEAEKPRNEEKYKQAKAEVSKKAEATLEKIKTDRKAAEBAARKAAE	400
Qy	173	DKVKEKP 179	
Db	401	DKVKEKP 407	

## RESULT 9

ID	Q9KK18	PRELIMINARY;	PRT;	660 AA.
----	--------	--------------	------	---------

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=srf2;  
RA Iannelli F.; Osgioni M.R.; Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus *pspc* of  
RT *Streptococcus pneumoniae*.";  
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF154038; AA73810.1; -  
DR InterPro; IPR002479; GSP binding.  
DR InterPro; IPR005877; Gpos YSTRK.  
DR Pfam; PF01473; CM binding\_1; 8  
DR Pfam; PF04650; YSTRK signal; 1.  
DR TIGRFAMs; TIGR01168; YSTRK signal; 1.  
SQ SEQUENCE 660 AA; 74515 MW; 374EB552FD09A659 CRC64;

SQ SEQUENCE 660 AA; 74515 MW; 374EB582FD09A659 CRC64

Query Match	Score	DB 2;	Length
91.0%	809;	DB 2;	Length 660;

Matches 172; Conservative 1; Mismatches 6; Indels 10; Gaps 2.

Qy	IKTDREAEAEAKRRDAK-----EOGPKGRARGVGELAPPDKKENAKSDS	51
Db	227 IKTDREKA-EAKRRADAKLTQENAVATSDQSKRRANREVLGELATPDDKENAKSDS	285
Qy	52 SVEEETLPSPSLKPEKVAEAKKVEEAKKAEDQKEEDRRNYPTNTYKTLLEIAESDV	111
Db	286 SVEEETLPSPSLKPEKVAEAKKVEEAKKAEDQKEEDRRNYPTNTYKTLLEIAESDV	345
Qy	112 EVKKAELVLVEBANEPRNEEKVAKAEVSSKKAELTLEKIKITDRKKAEEEEKKRAAE	171
Db	346 EVKKAELVLVEBANEPRNEEKVAKAEVSSKKAELTLEKIKITDRKKAEEEEKKRAAE	405
Qy	172 EDKVKEKPA 180	
Db	406 EDKVKEKPA 414	

## RESULT 10

ID Q9KK47 PRELIMINARY; PRT; 730 AA..

DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-Mar-2003 (Tremblrel. 23, Last annotation update)  
DE Surface protein PspC.  
GN PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=g363;  
RC Iannelli F., Ogiomi M.R., Pozzi G.;  
RA "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae";  
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AF154013, AAF73780.1; -  
DR InterPro; IPR002479; CW\_binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW\_binding\_1; 1.  
DR Pfam; PF04650; YsIRK\_signal; 1.  
DR TIGRFAMs; TIGR01168; YsIRK\_signal; 1.  
SQ SEQUENCE 730 AA; 82304 MW; D5C5D0BE4BD74E CRC64;

Query Match	90.8%;	Score 807;	DB 2;	Length 730;
-------------	--------	------------	-------	-------------

Matches 167; Conservative 7; Mismatches 6; Indels 12; Gaps 1;

QY 1 IKDREAREAEKRRADAC-----EQQPGPRAGVGGELAPPDKENDAKS 48

Db 225 IKDRKRAEEAKRRRAEAKLKAENVAVNATSEQQKPGPRAGVGGQALPPDKENDAKS 284

QY 49 SDSGVEELTLPSPSLPEKRYAAEAKKYAEAKKAAEQKEDRRNNTNTNYKTLLEIAE 108

Db 285 SSSGVEEELTLPSPSLPEKRYAAEAKKYAEAKKAAEQKEDRRNNTNTNYKTLLEIAE 344

QY 109 SDVEYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 168  
DB 345 SDVYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 404  
QY 169 AAEEDKVEKPA 180  
DB 405 AAEEDKVEKPA 416

RESULT 11  
QYK36 PRELIMINARY; PRT; 695 AA.  
ID 09K36  
AC 09K36; 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
GN Surface protein PspC.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=g394;  
RA Lannelli F., Oggioni M.R., Pozzi G.;  
RT "Allelic Variation in the Highly Polymorphic Locus pspC of  
RT Streptococcus pneumoniae."  
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF154023; AAF73792.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW binding\_1; 8.  
DR Pfam; PF04650; YsIRK signal; 1.  
DR TIGRfams; TIGR01168; YsIRK signal; 1.  
SQ SEQUENCE 695 AA; 78361 MW; F8C84F08C4EFFFB4 CRC64;

Query Match 90.4%; Score 804; DB 2; Length 695;  
Best Local Similarity 91.3%; Pred. No. 1.7e-37;  
Matches 168; Conservative 3; Mismatches 9; Indels 4; Gaps 1;

QY 1 IKTDREAEAEKRRADAK-----KPKGRKRGVPGELATPDKKENDAKSSDS 56  
DB 265 IKTDREAEAEKRRADAKEDDESSKRRKVRKGDGEQATPDKKENDAKSSDSVGE 324  
QY 57 TLPSLSLKEPKKVAEAEKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 116  
DB 325 TLPSLSLKEPKKVAEAEKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 384  
QY 117 ELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 176  
DB 385 ELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 444  
QY 177 EKPA 180  
DB 445 EKPA 448

RESULT 12  
069188 PRELIMINARY; PRT; 564 AA.  
ID 069188  
AC 069188; 01-AUG-1998 (TREMblrel. 07, Created)  
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
GN C3-binding protein.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=CP1200;  
RA MEDLINE=20281293; PubMed=10820017;  
RA Cheng Q., Finkel D., Hostetter M.K.;  
RT "Novel purification scheme and functions for a C3-binding protein from  
RT Streptococcus pneumoniae."  
RL Biochemistry 39:5450-5457(2000).  
DR EMBL; AF067128; AAC17445.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW binding\_1; 4.  
DR Pfam; PF04650; YsIRK signal; 1.  
DR TIGRfams; TIGR01168; YsIRK signal; 1.  
SQ SEQUENCE 564 AA; 63596 MW; 884C97BA04AED20 CRC64;

Query Match 90.4%; Score 803.5; DB 2; Length 564;  
Best Local Similarity 86.8%; Pred. No. 1.5e-37;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEKRRADAK-----EOGKPKGRKRGVPGELATPDKKENDAKSSDS 51  
DB 226 IKTDREAEAEKRRADAKLEAVNATSDQKPKGRKRGVPGELATPDKKENDAKSSDS 285  
QY 52 SVGEETLPSLSLKEPKKVAEAEKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 111  
DB 286 SVGEETLPSLSLKGKVAEAEKVEAEKKAQKQKEEDRRNYPNTYTKTLELEIAESDV 345  
QY 112 EYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 171  
DB 346 KYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 405  
QY 172 EDKVEKPA 180  
DB 406 EDKVEKPA 414

RESULT 13  
033742 PRELIMINARY; PRT; 581 AA.  
ID 033742  
AC 033742; 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
GN SpsA protein.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=type 47;  
RX MEDLINE=98010350; PubMed=9350867;  
RA Hammerschmidt S., Talay S.R., Brandtzaeg P., Chhatwal G.S.;  
RT "SpsA, a novel pneumococcal surface protein with specific binding to  
RT Immunoglobulin A and secretory component."  
RL Mol. Microbiol. 25:1113-1124(1997).  
DR EMBL; AU002055; CAA05159.1; -  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR005877; Gpos\_YsIRK.  
DR Pfam; PF01473; CW binding\_1; 4.  
DR Pfam; PF04650; YsIRK signal; 1.  
DR TIGRfams; TIGR01168; YsIRK signal; 1.  
SQ SEQUENCE 581 AA; 65671 MW; 9CF32F2BE4490DFC CRC64;

Query Match 90.4%; Score 803.5; DB 2; Length 581;  
Best Local Similarity 86.8%; Pred. No. 1.5e-37;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEKRRADAK-----EOGKPKGRKRGVPGELATPDKKENDAKSSDS 51  
DB 226 IKTDREAEAEKRRADAKLEAVNATSDQKPKGRKRGVPGELATPDKKENDAKSSDS 285  
QY 52 SVGEETLPSLSLKEPKKVAEAEKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 111

```
DB 286 SVGEETLPSSSLKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLDLLEIAESDV 345
QY 112 EVKKAELVLVKEAEKPEPNEEKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAE 171
DB 346 KVKAELELVKEAEKPEPDEEKIKQAKAKVESKKAATRLLENIKTDKKAEEBAKRAAE 405
QY 172 EDKVKKEPA 180
DB 406 EDKVKKEPA 414
```

## RESULT 14

```
030874 PRELIMINARY; PRT; 663 AA.
AC 030874;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Choline binding protein A (Fragment).
GN CBPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R6x;
RA Roenow C., Ryan P., Weiser J.N., Johnson S., Fontan P., Ortqvist A.,
RA Maure H.R.;
RT "Contribution of novel choline-binding proteins to adherence,
RT colonization and immunogenicity of Streptococcus pneumoniae.";
RL Mol. Microbiol. 0:0-0(1997).
DR EMBL; AF019904; AAB70838.1; -.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 10.
FT NON TER
SQ SEQUENCE 663 AA; 75064 MW; AFD825A7AB2ACE7B CRC64;
```

Query Match 90.4%; Score 803.5; DB 2; Length 663;  
Best Local Similarity 86.8%; Pred. No. 1.7e-37;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

```
QY 1 IKTDEEAEEBAKRRADAK-----EQQKPKGRAGVGPGLATPDKKENDAKSSDS 51
DB 188 IKTDRKKAEEBAKRRADAKLKEANVATSDQKPKGRAGVGPGLATPDKKENDAKSSDS 247
QY 52 SVGEETLPSSSLKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLLEIAESDV 111
DB 248 SVGEETLPSSSLKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLLEIAESDV 307
QY 112 EVKKAELVLVKEAEKPEPNEEKVKQAKAEVSKKAATRLKIKTDKKAEEBAKRAAE 171
DB 308 KVKAELELVKEAEKPEPDEEKIKQAKAKVESKKAATRLLENIKTDKKAEEBAKRAAE 367
QY 172 EDKVKKEPA 180
DB 368 EDKVKKEPA 376
```

## RESULT 15

```
09KK48 PRELIMINARY; PRT; 701 AA.
AC 09KK48;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1313;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D39;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154012; AAF73779.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 701 AA; 79098 MW; A80E31FE2846F1BF CRC64;
```

Query Match 90.4%; Score 803.5; DB 2; Length 701;  
Best Local Similarity 86.8%; Pred. No. 1.8e-37;  
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

```
QY 1 IKTDEEAEEBAKRRADAK-----EQQKPKGRAGVGPGLATPDKKENDAKSSDS 51
DB 226 IKTDRKKAEEBAKRRADAKLKEANVATSDQKPKGRAGVGPGLATPDKKENDAKSSDS 285
QY 52 SVGEETLPSSSLKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLLEIAESDV 111
DB 286 SVGEETLPSSSLKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLLEIAESDV 345
QY 112 EVKKAELVLVKEAEKPEPNEEKVKQAKAEVSKKAATRLKIKTDKKAEEBAKRAAE 171
DB 346 KVKAELELVKEAEKPEPDEEKIKQAKAKVESKKAATRLLENIKTDKKAEEBAKRAAE 405
QY 172 EDKVKKEPA 180
DB 406 EDKVKKEPA 414
```

Search completed: November 21, 2003, 13:33:49  
Job time : 16.6715 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 21.4457 Seconds  
(Without alignments)  
1008.166 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFASKSERKRVHYIRKFSVG.....NRLTQQQPKTEKPAQPSRP 511

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194.5	86.5	446	4	US-09-286-981B-9 Sequence 9, Appl
2	2177.5	85.8	446	4	US-09-286-981B-6 Sequence 6, Appl
3	1728	68.1	564	3	US-09-308-022-6 Sequence 6, Appl
4	1512.5	59.6	406	4	US-09-286-981B-18 Sequence 18, Appl
5	1506	59.4	631	3	US-08-847-065-25 Sequence 25, Appl
6	1455.5	57.4	426	4	US-09-286-981B-12 Sequence 12, Appl
7	1447.5	57.1	425	4	US-09-286-981B-13 Sequence 13, Appl
8	1446.5	57.0	424	4	US-09-286-981B-14 Sequence 14, Appl
9	1440.5	56.8	428	4	US-09-286-981B-7 Sequence 7, Appl
10	1432	56.4	425	4	US-09-286-981B-11 Sequence 11, Appl
11	1412.5	55.7	412	4	US-09-286-981B-17 Sequence 17, Appl
12	1410.5	55.6	419	4	US-09-286-981B-15 Sequence 15, Appl
13	1405.5	55.4	414	4	US-09-286-981B-16 Sequence 16, Appl
14	1401.5	55.2	414	4	US-09-286-981B-10 Sequence 10, Appl
15	1395	55.0	431	4	US-09-286-981B-3 Sequence 3, Appl
16	1373	54.1	453	3	US-08-961-083-38 Sequence 38, Appl
17	1373	54.1	453	4	US-09-536-784-38 Sequence 38, Appl
18	1291	50.9	413	4	US-09-286-981B-5 Sequence 5, Appl
19	1207	47.6	864	4	US-08-714-741-40 Sequence 40, Appl
20	1157	45.6	588	4	US-08-714-741-42 Sequence 42, Appl
21	904	35.6	605	4	US-08-714-741-46 Sequence 46, Appl
22	856.5	33.8	8991	4	US-08-714-741-32 Sequence 32, Appl
23	764	30.1	1231	4	US-08-714-741-41 Sequence 41, Appl
24	752	29.6	623	4	US-08-714-741-47 Sequence 47, Appl
25	714.5	28.2	251	4	US-09-286-981B-4 Sequence 4, Appl
26	636	25.1	142	3	US-08-847-065-21 Sequence 21, Appl
27	560	22.1	219	4	US-09-286-981B-8 Sequence 8, Appl

28	542	21.4	114	4	US-09-286-981B-19 Sequence 19, Appl
29	525.5	20.7	110	4	US-08-961-083-102 Sequence 102, App
30	525.5	20.7	110	3	US-09-536-784-102 Sequence 102, App
31	499.5	19.7	695	1	US-08-127-499A-23 Sequence 23, Appl
32	499.5	19.7	695	1	US-08-482-847-23 Sequence 23, Appl
33	491	19.4	619	1	US-08-465-746-2 Sequence 2, Appl
34	491	19.4	619	1	US-08-214-164-2 Sequence 2, Appl
35	491	19.4	619	2	US-08-467-852A-3 Sequence 3, Appl
36	491	19.4	619	2	US-08-246-616-2 Sequence 2, Appl
37	491	19.4	619	2	US-08-247-491A-3 Sequence 3, Appl
38	491	19.4	619	2	US-08-319-795-2 Sequence 2, Appl
39	491	19.4	619	3	US-08-468-985-2 Sequence 2, Appl
40	491	19.4	619	3	US-08-312-949-2 Sequence 2, Appl
41	491	19.4	648	1	US-08-072-070-2 Sequence 2, Appl
42	491	19.4	648	1	US-08-468-434-2 Sequence 2, Appl
43	491	19.4	648	1	US-08-214-222-2 Sequence 2, Appl
44	491	19.4	648	2	US-08-467-852A-2 Sequence 2, Appl
45	491	19.4	648	2	US-08-468-718-2 Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-286-981B-9

Sequence 9, Application US/09286981B

Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

CURRENT FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patent Ver. 2.1

SEQ ID NO 9

LENGTH: 446

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-9

Query Match 86.5%; Score 2194.5; DB 4; Length 446;  
Best Local Similarity 99.8%; Pred. No. 7.3e-146;  
Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	37	TENEGATQVPTSSNRANSOAEQGHOPKLDSEDPKAEVEYVKTI VGSYAKTKR	96
DB	1	TENEGATQVPTSSNRANSOAEQGHOPKLDSEDPKAEVEYVKTI VGSYAKTKR	60
QY	97	HTTATVAVNEENNKIKNEYLINKI VESTSESQOILMMESRSKYDEAVSFKFKDSSSSSSSD	156
DB	61	HTTATVAVNEENNKIKNEYLINKI VESTSESQOILMMESRSKYDEAVSFKFKDSSSSSSSD	120
QY	157	SSTPEASDTAKPKNPTPEGKVAEAKKVEEA-KKAKDQKEEDRRNPTTYTKLELEI	215
DB	121	SSTPEASDTAKPKNPTPEGKVAEAKKVEEA-KKAKDQKEEDRRNPTTYTKLELEI	180
QY	216	AESDVEVKKALELVKVAANPEDEOKIKOAEVSESQOAEATRLKIKITREAEFEAK	275
DB	181	AESDVEVKKALELVKVAANPEDEOKIKOAEVSESQOAEATRLKIKITREAEFEAK	240
QY	276	RRADAKEQKPKRAKGVPEGLATPDKENDAKSSSVGEETLPSPSLKPEKKVAEAE	335
DB	241	RRADAKEQKPKRAKGVPEGLATPDKENDAKSSSVGEETLPSPSLKPEKKVAEAE	300
QY	336	KKVEBAKKAEADQKEEDRRNPTTYTKLELEI AESSDVEVKKALELVKKEBAKPRNEEK	395
DB	301	KKVEBAKKAEADQKEEDRRNPTTYTKLELEI AESSDVEVKKALELVKKEBAKPRNEEK	360

QY 396 VKQAKAEVESKKAELTRLEKIKTDKRAKAEBAKKAABEDVKVEKPAEOPAPAPAKAEK 455  
 DB 361 VKQAKAEVESKKAELTRLEKIKTDKRAKAEBAKKAABEDVKVEKPAEOPAPAPAKAEK 420  
 QY 456 PAPAPKRENPAPAEOPKAEEKPADQAAE 481  
 DB 421 PAPAPKRENPAPAEOPKAEEKPADQAAE 446

RESULT 2  
 US-09-286-981B-6  
 ; Sequence 6, Application US/09286981B  
 ; Patent No. 6503511  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wizenmann, Theresa M.  
 ; APPLICANT: Koenig, Scott  
 ; APPLICANT: Johnson, Leslie S  
 ; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
 ; PRIORITY REFERENCE: 469201-396  
 ; CURRENT APPLICATION NUMBER: US/09/286,981B  
 ; PRIOR FILING DATE: 1999-04-06  
 ; PRIOR APPLICATION NUMBER: US 60/085,743  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 446  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 ; US-09-286-981B-6

Query Match 85.8%; Score 2177.5; DB 4; Length 446;  
 Best Local Similarity 98.9%; Pred. No. 1,1e-144;  
 Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;  
 QY 37 TENEGATQVPTSSNRANESQAEQGEOPKLDSEBDKARKKEVEEVYKKIVGESYAKSTKKR 96  
 DB 1 TENEGATQVPTSSNRANESQAEQGEOPKLDSEBDKARKKEVEEVYKKIVGESYAKSTKKR 60  
 QY 97 HTTVALVNLNNIKNEYLNKIVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSSD 156  
 DB 61 HTTVALVNLNNIKNEYLNKIVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSSD 120  
 QY 157 SSTKPEASDPAKNNKPTPEGEKVAEAKKVEEVEKAKDQKEDRRNYPTIYKTLLELEI 215  
 DB 121 SSTKPEASDPAKNNKPTPEGEKVAEAKKVEEVEKAKDQKEDRRNYPTIYKTLLELEI 180  
 QY 216 AESDVEVKAELVYKVKANEPDEOKIKQAEAEVESKQAEATRLKIKITDREAEAEAK 275  
 DB 181 AESDVEVKAELVYKVKANEPDEOKIKQAEAEVESKQAEATRLKIKITDREAEAEAK 240  
 QY 276 RRADAEQCKPKGRKRGVDELATPDCKENDAKSSDSSVGEETLPSPLKPEKVAEAE 335  
 DB 241 RRADAEQCKPKGRKRGVDELATPDCKENDAKSSDSSVGEETLPSPLKPEKVAEAE 300  
 QY 336 KKYVEAKKKAEDQKEDRRNYPTNTYKTLLEIAESVVEVKAELVYKVEAEKPEPNEEK 395  
 DB 301 KKYVEAKKKAEDQKEDRRNYPTNTYKTLLEIAESVVEVKAELVYKVEAEKPEPNEEK 360  
 QY 396 VKQAKAEVESKKAELTRLEKIKTDKRAKAEBAKKAABEDVKVEKPAEOPAPAPAKAEK 455  
 DB 361 VKQAKAEVESKKAELTRLEKIKTDKRAKAEBAKKAABEDVKVEKPAEOPAPAPAKAEK 420  
 QY 456 PAPAPKRENPAPAEOPKAEEKPADQAAE 481  
 DB 421 PAPAPKRENPAPAEOPKAEEKPADQAAE 446

RESULT 3  
 US-09-308-022-6  
 ; Sequence 6, Application US/09308022  
 ; Patent No. 6291654  
 ; GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.  
 TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS  
 TITLE OF INVENTION: PNEUMONIAE  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MEETING, RAASCH & GERHARDT, P.A.  
 STREET: 119 No. 6291654th Fourth Street, Suite 203  
 CITY: Minneapolis  
 STATE: Minnesota  
 COUNTRY: USA  
 ZIP: 55401  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/308,022  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US97/20586  
 FILING DATE: 12-NOV-1997  
 APPLICATION NUMBER: 60/029,444  
 FILING DATE: 12-NOV-1996  
 APPLICATION NUMBER: 60/038,086  
 FILING DATE: 18-FEB-1997  
 APPLICATION NUMBER: 60/059,368  
 FILING DATE: 19-SEP-1997  
 APPLICATION NUMBER: 60/062,473  
 FILING DATE: 16-OCT-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MOETING, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 110 00430101  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-305-1217  
 TELEFAX: 612-305-1228  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 564 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-308-022-6

Query Match 68.1%; Score 1728; DB 3; Length 564;  
 Best Local Similarity 70.5%; Pred. No. 3.8e-113;  
 Matches 370; Conservative 50; Mismatches 47; Indels 58; Gaps 10;  
 QY 1 MPASKSEKRVHYSIRKFSVG-ASVVVASLWGSVVAHATENEGATQVPTSSNRANESQAEQ 59  
 DB 1 MPASKSEKRVHYSIRKFSVG-ASVVVASLWGSVVAHATENEGATQVPTSSNRANESQAEQ 54  
 QY 60 GQOPKLDSEBDKARKKEVEEVYKKIVGESYAKSTKKRTIYKTLLEIAESVVEVKAELVYKVKAN 117  
 DB 55 -----TEHRAKAKQVVDYIEKMLRE--IOLDRKRTQVVALNLIKSAIKTKYLR 104  
 QY 118 -IVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSDSTKPEASDPAKNNKPTPEB 176  
 DB 105 NVLEESKDEL---SEIKAKIDAAFEKTKOT-----LKG 138  
 QY 177 EKVAAEKKKVVEA-KKAKDQKEDRRNYPTIYKTLLEIAESVVEVKAELVYKVKAN 235  
 DB 139 EKVAAEKKKVVEA-KKAKDQKEDRRNYPTIYKTLLEIAESVVEVKAELVYKVKAN 198  
 QY 236 EPRDEOKIKQAEAEVSKQAEATRLKIKITDREAEAEKKAADAK-----EQGRP 286  
 DB 199 ESRNEGTTIKQAEKESKKAELTRLEKIKTDKRAKAEBAKKAABEDVKVEKPAEOPAPAPAKAEK 258  
 QY 287 KGRKRGVDELATPDCKENDAKSSDSSVGEETLPSPLKPEKVAEAEKVEAEKKAAB 346

Db 259 KGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSSSLKSGKVAEAKKEVEAEKKA 318  
Qy 347 DQKEDERNPTNTYKTLLEIAESDVVKAELELVKEAKPRNEEKKOKAEVSK 406  
Db 319 DQKEDERNPTNTYKTLLEIAESDVVKAELELVKEAKPRNEEKKOKAEVSK 378  
Qy 407 KAENTRLEKIKTRKKAEEBAKKAEDVYKEKPAEOPAPAPAKPAKPENPA 466  
Db 379 KAENTRLEKIKTRKKAEEBAKKAEDVYKEKPAEOPAPAPAKPAKPENPA 436  
Qy 467 EQPKAEKPADQAEEDYARSRSEERYNLTQOOPKTEKPAOPSTP 511  
Db 437 EQPKAEKPADQAEEDYARSRSEERYNLTQOOPKTEKPAOPSTP 481

RESULT 4  
US-09-286-981B-18  
Sequence 18, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizenmann, Theresa M.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 18  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-18

Query Match 59.6%; Score 1512.5; DB 4; Length 406;  
Best Local Similarity 73.0%; Pred. No. 2.9e-98;

Matches 327; Conservative 28; Mismatches 46; Indels 47; Gaps 8;

Qy 37 TENEGATOVPTSSNRANESQAEQEQPKLDSERDARKEVEEYKIKVGEASYAKSTKR 96  
Db 1 TENEGTTOAPTSSNRGNEQA-----EHMKAQVDEYIEKML-----QDDRK 44  
Qy 97 HTTVLALVNEINNIKNEYLANKIVESTESQOLIMMESKSKVDNAVSKTEKSSSSSSD 156  
Db 45 HTQNVGLTLTGAIKTEYLRLGLSVSKKSTAE-LPSEIKKLTAAFKQPKK----- 94  
Qy 157 SSTKPEASPTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEDERNPTNTYKTLLEI 215  
Db 95 -----DTLAKK-----KVAEAKKVAEAKKAEEDQKEDERNPTNTYKTLLEI 140  
Qy 216 AESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEAEK 275  
Db 141 AESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEAEK 200  
Qy 276 RRDADAEQ-----KPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKV 331  
Db 201 RRDADAEQ-----KPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKV 260  
Qy 332 AEAKEKVEAEKKAEDQKEDERNPTNTYKTLLEIAESDVVKAELELVKEAKEPR 391  
Db 261 AEAKEKVEAEKKAEDQKEDERNPTNTYKTLLEIAESDVVKAELELVKEAKEPR 320  
Qy 392 NEKVQKAEAEVSKQAEATRLKKITDRKKAEEBAKKAEDVYKEKPAEOPAPAP 451  
Db 321 NEKVQKAEAEVSKQAEATRLKKITDRKKAEEBAKKAEDVYKEKPAEOPAPAP 380  
Qy 452 KAEKPAKPAKPENPAOPKAEKPADQQA 479  
Db 381 QPKK--PAKPPENPAOPKAEKPADQQA 406

RESULT 5  
US-08-847-065-25  
Sequence 25, Application US/08847065  
Patent No. 6245335  
GENERAL INFORMATION:  
APPLICANT: Masure, H. Robert  
APPLICANT: Rosenow, Carsten I.  
APPLICANT: Tuomanen, Elaine  
APPLICANT: Wizenmann, Theresa M.  
TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/847,065  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-158  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-1684  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 631 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-847-065-25

Query Match 59.4%; Score 1506; DB 3; Length 631;  
Best Local Similarity 73.5%; Pred. No. 1.4e-97;  
Matches 316; Conservative 39; Mismatches 35; Indels 40; Gaps 6;

Qy 94 KKRHTTVLALVNEINNIKNEYLANK--IVESTESQOLIMMESKSKVDNAVSKPEKSSS 151  
Db 9 RKRHTQNALVNIKLSAITKTYLRLNLVLEKSKDEL--SEIRKXKLAAPFKKDT-- 63  
Qy 152 SSSSSSTKPEASPTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEDERNPTNTYKT 210  
Db 64 -----LKPGEKVAEAKKVEEA-KKAKQKEDERNPTNTYKT 102  
Qy 211 LELEIAESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEA 270  
Db 103 LELEIAESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRKA 162  
Qy 271 EEAARRADAK-----EQPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLP 321  
Db 163 EEAARRADAKLANVATSDQPKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLP 222  
Qy 322 SPSPKPEKVAEAKKVEAEKKAEDQKEDERNPTNTYKTLLEIAESDVVKAELE 381  
Db 223 SSSSLKSGKVAEAKKVEAEKKAEDQKEDERNPTNTYKTLLEIAESDVVKAELE 282  
Qy 382 LVKEAKEPRNEKVQKAEAEVSKQAEATRLKKITDRKKAEEBAKKAEDVYKEKPA 441

Db 283 LVEAEKPEDEEKTQAKAKVSKAEATRLNKTDRKKAEEBAKRAKAEEDKVEKXP 342  
Qy 442 AEOPAPAPAKAEKAPAPKPPENPAOPKAEKPADQOAEEDVYARSEENRNLTOOQPPK 501  
Db 343 AEOPAPAPAPQPEK-PAPPEKPAEOPKAEKTDQOAEEDVYARSEENRNLTOOQPPK 400  
Qy 502 TEKPAOPSTP 511  
Db 401 TEKPAOPSTP 410

RESULT 6  
US-09-286-981B-12  
Sequence 12, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 12  
LENGTH: 426  
TYPE: PR1  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-12

Query Match 57.4%; Score 1455.5; DB 4; Length 426;  
Best Local Similarity 68.4%; Pred. No. 3e-94;  
Matches 318; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

Qy 37 TENEGATQVPTSSNRANESQAEQGEOPKLDSEEDKARKEVEEYKIVGESYAKSTKR 96  
Db 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAAQVDEYIKKKI-----QIDRRK 44  
Qy 97 HTTVALVNLNNKNEYLNKIIVESTESQOILMESRSKYDEAVSKFEKSSSSSSD 156  
Db 45 HTQNVGLTLKGLVITKTEYHLGSLVSKKSEAE-LPSEIKAKLDAAFEQPKDT----- 96  
Qy 157 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDOKEEDRRNYPTIYKTLLEI 215  
Db 97 -----LPTEPGKVAEAKKVEEAKKKAEDQEKDRLNYPNTYKTLLELDI 142  
Qy 216 AESDVEYKAELELVYKVAKEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEBAK 275  
Db 143 AESDVEYKAELELVKEEAKESRDEKINQAKAVENKKAETRLKIKITDREKA-EBAK 201  
Qy 276 RRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDSSVGEETLTPSLK 326  
Db 202 RRADAKQAEANVATSEODKSRRAKREVLGELATPDKKENDAKSSDSSVGEETLTPSLK 261  
Qy 327 PEKKVAEAKKVEEAKKKAEDQEKEDRRNYPTNTYKTLLEIAESDVEYKAELELVKEE 386  
Db 262 PEKKVAEAKKVEEAKKKAEDQEKEDRRNYPTNTYKTLLEIAESDVEYKAELELVKEE 321  
Qy 387 AKEPRNEKYKQAEVESKKAETRLKIKITDRKKA-EBAKRAAEEDKVEKPAEOP 445  
Db 322 AKESRNEEKIKQVAKAVESKKAETRLKIKITDRKKAEEBAKRAAEEDKVEKPAEOP 381  
Qy 446 QPAPAPAKAEK-----PAPKPPENPAOPKAEKPADQOAE 481  
Db 382 QPAPAPQPEKTEEPENPAOPAPKPPENPAEKPAEKPADQOAE 426

RESULT 7  
US-09-286-981B-13

Sequence 13, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 13  
LENGTH: 425  
TYPE: PR1  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-13

Query Match 57.1%; Score 1447.5; DB 4; Length 425;  
Best Local Similarity 68.3%; Pred. No. 1.1e-93;  
Matches 317; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

Qy 37 TENEGATQVPTSSNRANESQAEQGEOPKLDSEEDKARKEVEEYKIVGESYAKSTKR 96  
Db 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAAQVDEYIKKKI-----QIDRRK 44  
Qy 97 HTTVALVNLNNKNEYLNKIIVESTESQOILMESRSKYDEAVSKFEKSSSSSSD 156  
Db 45 HTQNVGLTLKGLVITKTEYHLGSLVSKKSEAE-LPSEIKAKLDAAFEQPKDT----- 96  
Qy 157 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDOKEEDRRNYPTIYKTLLEI 215  
Db 97 -----LPTEPGKVAEAKKVEEAKKKAEDQEKDRLNYPNTYKTLLELDI 142  
Qy 216 AESDVEYKAELELVYKVAKEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEBAK 275  
Db 143 AESDVEYKAELELVKEEAKESRDEKINQAKAVENKKAETRLKIKITDREKA-EBAK 201  
Qy 276 RRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDSSVGEETLTPSLK 326  
Db 202 RRADAKQAEANVATSEODKSRRAKREVLGELATPDKKENDAKSSDSSVGEETLTPSLK 261  
Qy 327 PEKKVAEAKKVEEAKKKAEDQEKEDRRNYPTNTYKTLLEIAESDVEYKAELELVKEE 386  
Db 262 PEKKVAEAKKVEEAKKKAEDQEKEDRRNYPTNTYKTLLEIAESDVEYKAELELVKEE 321  
Qy 387 AKEPRNEKYKQAEVESKKAETRLKIKITDRKKA-EBAKRAAEEDKVEKPAEOP 445  
Db 322 AKESRNEEKIKQVAKAVESKKAETRLKIKITDRKKAEEBAKRAAEEDKVEKPAEOP 381  
Qy 446 QPAPAPAKAEK-----PAPKPPENPAOPKAEKPADQOAE 480  
Db 382 QPAPAPQPEKTEEPENPAOPAPKPPENPAEKPAEKPADQOAE 425

RESULT 8  
US-09-286-981B-14  
Sequence 14, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 14  
LENGTH: 424  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-14

Query Match 57.0%; Score 1446.5; DB 4; Length 424;  
Best Local Similarity 68.5%; Pred. No. 1.2e-93;  
Matches 317; Conservative 40; Mismatches 47; Indels 59; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQGPGLDSEPDKARKEVEEYKKTIGESYAKSTKR 96  
DB 1 TEKEVTTQVATSSNRANESQAEQGPGLDSEPDKARKEVEEYKKTIGESYAKSTKR 44  
QY 97 HTITVALVNEIINNKNKYLINKVESTSESQOLIMMESRSKVDPAVSKPEKSSSSSSD 156  
DB 45 HTQVGLTITKLVITKTYLHGLSVSKSSSEAE-LPSEIKAKLDAEFOFKDT----- 96  
QY 157 SSTPEASDTAKPKPTPEGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEI 215  
DB 97 -----LPTEPGKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEI 142  
QY 216 AESDVEVKAELEIVKANEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEAEAK 275  
DB 143 AESDVEVKAELEIVKEAEKESRDEKINOAKAVENKKAATRLKIKITDREKA-EAK 201  
QY 276 RRADAK-----EQGPKGRAGVGPGLATPDKKENDAKSSSVGEETLPSLSLK 326  
DB 202 RRADAKQOEANVATSEODKRRKREVLELATPDKKENDAKSSSVGEETLPSLSLK 261  
QY 327 PEKKVAEAEKKEAEKKAEDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVKEE 386  
DB 262 PEKKVAEAEKKEAEKKAEDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVKEE 321  
QY 387 AKERNEEYKQAKAEVSKAEATRLKIKITDRKKA-BEAKRKAEDKVEKPAEOP 445  
DB 322 AKSRNEEKIKQYAKAVESKKAATRLKIKITDRKKA-BEAKRKAEDKVEKPAEOP 381  
QY 446 QPAPAPAEK-----PAPAPKPEPAPAEKPAEOP 479  
DB 382 QPAPAPAEKPEPAPAEKPAEOP 424

RESULT 9

US-09-286-981B-7  
Sequence 7, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
PLICANT: Koenig, Scott  
PLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 7  
LENGTH: 428  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-7

Query Match 56.8%; Score 1440.5; DB 4; Length 428;  
Best Local Similarity 69.7%; Pred. No. 3.3e-93;  
Matches 311; Conservative 35; Mismatches 53; Indels 47; Gaps 7;

QY 60 GEQPKLDEBDKAR---EVEEYVKKIVGESYAKSTKRHTITVALVNEIINNKNKYLN 116  
DB 6 GNNSTVTSGGDISKYADEVESHQSLIKDVNKLKKTQHTQMDPNKKLSKITKTYLY 65

QY 117 K-IVESTESOLQILIMMESRSKVDPAVSKPEKSSSSSSSSTKPEASDTAKPKKPT 174  
DB 66 ELNVLEKSEBAELTSTKTEKTEELTAFOFKDTIS-----IT 104  
QY 175 PGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVK 233  
DB 105 PEKKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVK 164  
QY 234 ANPRDEOKIKQAEAVESKQAEATRLKIKITDREAE-----EAKRRADAE 282  
DB 165 ANPRDEOKIKQAEAVESKQAEATRLKIKITDREAEATRLKIKITDREAEAEKVK 224  
QY 283 QGPKGRAGVGPGLATPDKKENDAKSSSVGEETLPSLSLKPEKVAEAKKVEEAK 342  
DB 225 E-PKRTKRGVLGEATPDKKENDAKSSSVGEETLPSLSLKPEKVAEAKKVEEAK 282  
QY 343 KKAEDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVKEAEKPRNEEKYQAKAE 402  
DB 283 KKAEDQKEDRRNYPITTYKLELEIAESDVEVKAELEIVKEAEKPRNEEKYQAKAE 342  
QY 403 VESKKAATRLKIKITDRKKAEEAKRKAEDKVEKPAEOPAPAEKPAEOP 455  
DB 343 VESKKAATRLKIKITDRKKAEEAKRKAEDKVEKPAEOPAPAEKPAEOP 402  
QY 456 PAPAPKPEPAPAEKPAEOP 481  
DB 403 PAPAPKPEPAPAEKPAEOP 428

RESULT 10

US-09-286-981B-11  
Sequence 11, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wizemann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 11  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-11

Query Match 56.4%; Score 1432; DB 4; Length 425;  
Best Local Similarity 66.6%; Pred. No. 1.3e-92;  
Matches 311; Conservative 44; Mismatches 48; Indels 64; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQGPGLDSEPDKARKEVEEYKKTIGESYAKSTKR 96  
DB 1 TEKEVTTQVATSSNRANESQAEQGPGLDSEPDKARKEVEEYKKTIGESYAKSTKR 44  
QY 97 HTITVALVNEIINNKNKYLINK-IVESTESOLQILIMMESRSKVDPAVSKPEKSSSSSS 154  
DB 45 HTQVGLTITKLVITKTYLHGLSVSKSSSEAE-LPSEIKAKLDAEFOFKDT----- 101  
QY 155 SDSSTPEASDTAKPKPTPEGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEI 213  
DB 102 T-----VAEAEKVEEAKKKAQKEDRRNYPITTYKLELEI 138  
QY 214 EIASDVEVKAELEIVKANEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEAE 273  
DB 139 EIASDVEVKAELEIVKEAEKESRDEKIKQAEAVESKQAEATRLKIKITDREAEAE 198  
QY 274 AKRRADAK-----EQGPKGRAGVGPGLATPDKKENDAKSSSVGEETLPSLSLK 321



PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 16  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-16

Query Match 55.4%; Score 1405.5; DB 4; Length 414;  
Best Local Similarity 67.7%; Pred. No. 8,9e-91;  
Matches 310; Conservative 45; Mismatches 46; Indels 57; Gaps 11;

QY 37 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 95  
DB 1 TENEGSTQATSSNMAK-----TEHRAAKQVDEVIEKMLRE--IQLDR 44  
QY 96 RHITVALVNLNNIKNEYLNK--IVESTSESQLOIMMERSKVDKAVSKFEKSSSS 153  
DB 45 KHTQVVALNLIKLSAIKTKYLRNLVLEEKSKDELPE--SEIKAKLDAA--FEK----- 92  
QY 154 SSSSTPEASDTAKPKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLE 212  
DB 93 -----EKKDILK-----PGEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLE 138  
QY 213 LEIASDVEVKKALELVKVKANEPDEQIKQAEAVESKQAEATRLKIKITDREAE 272  
DB 139 LEIAEDVVKKALELVKVEAKESRNEGTIKQAEKVESKKAENRLNIKTRKKAEE 198  
QY 273 EAKRADAK-----EQQPKRAKRGVPELATPDKKNDKSSSVGEETLPS 323  
DB 199 EAKRKADAKLKEANVATSDQKPKRAKRGVPELATPDKKNDKSSSVGEETLPS 258  
QY 324 SLKPEKKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEVKKALELV 383  
DB 259 SLKSGKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEVKKALELV 318  
QY 384 KEAKEPRNEEKVKQAKAEVSKKAATRLKIKITDRKKAEEAKKAEEDVKKEPAE 443  
DB 319 KEAKEPRDEEIKQAKAKVESKKAATRLNIKTRKQAEAKKAEEDVKKEPAE 378  
QY 444 QOPAPAPAKPAKPAKPEPAPKPAKPADQAAE 481  
DB 379 QOPAPATQPEK--PAPKPEKPAEOPKAEKTDQAAE 414

RESULT 14  
US-09-286-981B-10  
Sequence 10, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 10  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-10

Query Match 55.2%; Score 1401.5; DB 4; Length 414;  
Best Local Similarity 66.6%; Pred. No. 1.7e-90;  
Matches 305; Conservative 48; Mismatches 48; Indels 57; Gaps 9;  
QY 37 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 95

DB 1 TENEGSTQATSSNMAK-----TEHRAAKQVDEVIEKMLRE--IQLDR 44  
QY 96 RHITVALVNLNNIKNEYLNK--IVESTSESQLOIMMERSKVDKAVSKFEKSSSS 153  
DB 45 KHTQVVALNLIKLSAIKTKYLRNLVLEEKSKDELPE--SEIKAKLDAAFEKFKOT---- 97  
QY 154 SSSSTPEASDTAKPKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLE 212  
DB 98 -----LKEPEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLE 138  
QY 213 LEIASDVEVKKALELVKVKANEPDEQIKQAEAVESKQAEATRLKIKITDREAE 272  
DB 139 LEIAEDVVKKALELVKVEAKESRNEGTIKQAEKVESKKAENRLNIKTRKKAEE 198  
QY 273 EAKRADAK-----EQQPKRAKRGVPELATPDKKNDKSSSVGEETLPS 323  
DB 199 EAKRKADAKLKEANVATSDQKPKRAKRGVPELATPDKKNDKSSSVGEETLPS 258  
QY 324 SLKPEKKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEVKKALELV 383  
DB 259 SLKSGKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEVKKALELV 318  
QY 384 KEAKEPRNEEKVKQAKAEVSKKAATRLKIKITDRKKAEEAKKAEEDVKKEPAE 443  
DB 319 KEAKEPRDEEIKQAKAKVESKKAATRLNIKTRDRAEEAKKAEEDVKKEPAE 378  
QY 444 QOPAPAPAKPAKPAKPEPAPKPAKPADQAAE 481  
DB 379 QOPAPATQPEK--PAPKPEKPAEOPKAEKTDQAAE 414

RESULT 15  
US-09-286-981B-3  
Sequence 3, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-3

Query Match 55.0%; Score 1395; DB 4; Length 431;  
Best Local Similarity 64.0%; Pred. No. 5.1e-90;  
Matches 304; Conservative 57; Mismatches 60; Indels 54; Gaps 9;  
QY 37 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 96  
DB 1 TEKEVTTPTVATSSNMANKSQ-----TEHRAAKQVDEVIEKMLRE--IQLDR 44  
QY 97 HTITVALVNLNNIKNEYLNKIVESTSESQLOIMMERSKVDKAVSKFEKSSSSSD 156  
DB 45 HTQVVALNLIKLSAIKTKYLRNLVLEEKSKDELPELTSKKEIDAFQFNKDT----- 97  
QY 157 SSTRPEASDTAKPKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLEI 215  
DB 98 -----LKEPEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLEI 141  
QY 216 AESDVEVKKALELVKVKANEPDEQIKQAEAVESKQAEATRLKIKITDREAEAK 275  
DB 142 AESDVEVKKALELVKVEAKESRNEGTIKQAKAVESKKAATRLKIKITDRKKAEEAK 201

QY 276 RRADAKEQ--GKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK----- 326  
Db 202 RKAEEVEVKVKLKKRTKRGAFGEPAATPDKKENDAKSSDSSVVKKS-SKPIKSEKKVAE 260  
QY 327 PEKKVAEAEKKVEAEKAKKEDOKEDRRNYPTNTYKTLELEIAESDVEVKAELELVKEE 386  
Db 261 AEKKVAEAEKKVAEAEKAKDOKEDRRNYPTNTYKTLELEIAESDVKVAELELVKEE 320  
QY 387 AKEPRNEKVKOAKAEVESKKAETRLLEKIKTDRKKAEEBAKRAAEEDVKVEKPAEQPQ 446  
Db 321 AKEPQNEKIKQAKAKVESKKAETRLLEKIKTRKKA-EEAKRKVAEEDVKVEKPAEQPQ 379  
QY 447 PAPAPKAEPAPAPKFPENPAEQPKAEKPADQQAEDYARSSSEBYNRLTQOOPPK 501  
Db 380 PAPAP--KPAPAPQPEKPAEQPKAEKPADQQAEDYARSSSEBYNPLDLTAPAK 431

h completed: November 21, 2003, 13:35:10  
ime : 22.4457 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 7.55427 Seconds  
(without alignments)  
1008.166 Million cell updates/sec

Title: US-09-298-523C-13\_COPY\_263\_442  
Perfect score: 889  
Sequence: 1 IKTDREAEAEAKRRADAKE.....AEEBAKRAAEEDKVEKRA 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	446	4 US-09-286-981B-9	Sequence 9, Appl
2	889	100.0	453	3 US-08-961-083-38	Sequence 38, Appl
3	889	100.0	453	4 US-09-536-784-38	Sequence 38, Appl
4	884	99.4	446	4 US-09-286-981B-6	Sequence 6, Appl
5	803.5	90.4	414	4 US-09-286-981B-16	Sequence 16, Appl
6	803.5	90.4	631	3 US-08-847-065-25	Sequence 25, Appl
7	801.5	90.2	564	3 US-09-308-022-6	Sequence 6, Appl
8	792	89.1	406	4 US-09-286-981B-18	Sequence 18, Appl
9	792	89.1	425	4 US-09-286-981B-11	Sequence 11, Appl
10	791.5	89.0	414	4 US-09-286-981B-10	Sequence 10, Appl
11	784	88.2	428	4 US-09-286-981B-7	Sequence 7, Appl
12	775.5	87.2	412	4 US-09-286-981B-17	Sequence 17, Appl
13	770.5	86.7	424	4 US-09-286-981B-14	Sequence 14, Appl
14	770.5	86.7	426	4 US-09-286-981B-12	Sequence 12, Appl
15	769.5	86.6	425	4 US-09-286-981B-13	Sequence 13, Appl
16	768	86.4	419	4 US-09-286-981B-15	Sequence 15, Appl
17	717.5	80.7	413	4 US-09-286-981B-5	Sequence 5, Appl
18	684.5	77.0	864	4 US-08-714-741-40	Sequence 40, Appl
19	676.5	76.1	431	4 US-09-286-981B-3	Sequence 3, Appl
20	669	75.3	588	4 US-08-714-741-42	Sequence 42, Appl
21	557.5	62.7	219	4 US-09-286-981B-8	Sequence 8, Appl
22	542	61.0	114	4 US-09-286-981B-19	Sequence 19, Appl
23	521	58.6	142	3 US-08-847-065-21	Sequence 21, Appl
24	517	58.2	251	4 US-09-286-981B-4	Sequence 4, Appl
25	487	54.8	103	4 US-09-286-981B-1	Sequence 1, Appl
26	461.5	51.9	605	4 US-08-714-741-46	Sequence 46, Appl
27	451	50.7	103	4 US-09-286-981B-38	Sequence 38, Appl

28	447.5	50.3	1231	4 US-08-714-741-41	Sequence 41, Appl
29	394.5	44.4	623	4 US-08-714-741-47	Sequence 47, Appl
30	375.5	42.2	8991	4 US-08-714-741-32	Sequence 32, Appl
31	306	34.4	110	3 US-08-961-083-102	Sequence 102, Appl
32	306	34.4	110	4 US-09-536-784-102	Sequence 102, Appl
33	237.5	26.7	55	4 US-09-286-981B-37	Sequence 37, Appl
34	182.5	20.5	1507	3 US-08-929-329-5	Sequence 5, Appl
35	174	19.6	468	4 US-09-328-352-6321	Sequence 6321, Ap
36	159.5	17.9	611	4 US-09-216-393B-81	Sequence 81, Appl
37	158.5	17.8	288	3 US-08-312-949-4	Sequence 4, Appl
38	158.5	17.8	288	3 US-08-446-201-4	Sequence 4, Appl
39	158.5	17.8	619	1 US-08-465-746-2	Sequence 2, Appl
40	158.5	17.8	619	1 US-08-214-164-2	Sequence 2, Appl
41	158.5	17.8	619	2 US-08-467-852A-3	Sequence 3, Appl
42	158.5	17.8	619	2 US-08-246-636-2	Sequence 2, Appl
43	158.5	17.8	619	2 US-08-247-491A-3	Sequence 3, Appl
44	158.5	17.8	619	2 US-08-319-795-2	Sequence 2, Appl
45	158.5	17.8	619	2 US-08-468-985-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-286-981B-9
; Sequence 9, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizenmann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-336
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-9
Query Match      100.0%; Score 889; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.2e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREAEAEAKRRADAKEQCKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60
DB 228 IKTDREAEAEAKRRADAKEQCKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 287
QY 61 PSLEKEKVAEAEKVEAEKKAEDQCKEDRRNYPTNTYKTELEIAESDVYKAELEL 120
DB 268 PSLEKEKVAEAEKVEAEKKAEDQCKEDRRNYPTNTYKTELEIAESDVYKAELEL 347
QY 121 VKEAEKEPRNEKVKQAFAVESSKKAETRLKIKITDRKKAEEAKRKAEDVKYKEXPA 180
DB 348 VKEAEKEPRNEKVKQAFAVESSKKAETRLKIKITDRKKAEEAKRKAEDVKYKEXPA 407

RESULT 2
US-08-961-083-38
; Sequence 38, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

```

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-38

Query Match 100.0%; Score 889; DB 3; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1,3e-65;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60  
25 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 84  
61 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 120  
85 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 144  
121 VKEAEKPRNEEKVKAKAEVESKKAATRLKIKTRKKAEEBAKKAEEEDVKKEPA 180  
145 VKEAEKPRNEEKVKAKAEVESKKAATRLKIKTRKKAEEBAKKAEEEDVKKEPA 204

RESULT 3  
US-09-536-784-38  
Sequence 38, Application US/09536784  
Patent No. 6573082  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-09-536-784-38

Query Match 100.0%; Score 889; DB 4; Length 453;  
Best Local Similarity 100.0%; Pred. No. 1,3e-65;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60  
25 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 84  
61 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 120  
85 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 144  
121 VKEAEKPRNEEKVKAKAEVESKKAATRLKIKTRKKAEEBAKKAEEEDVKKEPA 180  
145 VKEAEKPRNEEKVKAKAEVESKKAATRLKIKTRKKAEEBAKKAEEEDVKKEPA 204

RESULT 4  
US-09-286-981B-6  
Sequence 6, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Cholera Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 446  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-6

Query Match 99.4%; Score 884; DB 4; Length 446;  
Best Local Similarity 99.4%; Pred. No. 3,2e-65;  
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60  
228 IKTDREAEAEERARADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 287  
61 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 120  
288 PSIKPEKVAEAKKVEAKKADQEDRRNYPNTYKTLLEIAESDVEVKAELEL 347  
121 VKEAEKPRNEEKVKAKAEVESKKAATRLKIKTRKKAEEBAKKAEEEDVKKEPA 180

Db 348 VKEEAKEPNEEKVKQAKAEVESKKAETRLLEKI KTDKKAEBAKKAEDDKVKEKPA 407

RESULT 5

US-09-286-981B-16

Sequence 16, Application US/09286981B

Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

CURRENT FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.7.1

SEQ ID NO 16

LENGTH: 414

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-16

Query Match

Best Local Similarity 90.4%; Score 803.5; DB 4; Length 414;

Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

Db 1 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVPGELATPPDKENDAKSSDS 51

Db 189 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVPGELATPPDKENDAKSSDS 248

Qy 52 SVGEETLPSLSLKPEKVAEAEKVEBAKKAADQEDRRNYPNTYKTLLEIAESDV 111

Db 249 SVGEETLPSLSLKPEKVAEAEKVEBAKKAADQEDRRNYPNTYKTLLEIAESDV 308

Qy 112 EVKKAELVLVEBAEPNEEKVKQAKAEVESKKAETRLLEKI KTDKKAEBAKKA 171

Db 309 KYKAELELVKEBAKPEPBDEKIKQAKAVESKKAETRLLEKI KTDKKAEBAKKA 368

Qy 172 EDKVEKPA 180

Db 369 EDKVEKPA 377

RESULT 6

US-08-847-065-25

Sequence 25, Application US/08847065

Patent No. 6245335

GENERAL INFORMATION:

APPLICANT: Masure, H. Robert

APPLICANT: Rosenow, Carsten I.

APPLICANT: Tuomaneen, Elaine

APPLICANT: Wizemann, Theresa M.

TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Floor

STATE: Hackensack

COUNTRY: New Jersey

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/847,065

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-158

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 631 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: N-terminal

US-08-847-065-25

Query Match

Best Local Similarity 90.4%; Score 803.5; DB 3; Length 631;

Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

Qy 1 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVPGELATPPDKENDAKSSDS 51

Db 155 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVPGELATPPDKENDAKSSDS 214

Qy 52 SVGEETLPSLSLKPEKVAEAEKVEBAKKAADQEDRRNYPNTYKTLLEIAESDV 111

Db 215 SVGEETLPSLSLKPEKVAEAEKVEBAKKAADQEDRRNYPNTYKTLLEIAESDV 274

Qy 112 EVKKAELVLVEBAEPNEEKVKQAKAEVESKKAETRLLEKI KTDKKAEBAKKA 171

Db 275 KYKAELELVKEBAKPEPBDEKIKQAKAVESKKAETRLLEKI KTDKKAEBAKKA 334

Qy 172 EDKVEKPA 180

Db 335 EDKVEKPA 343

RESULT 7

US-09-308-022-6

Sequence 6, Application US/09308022

Patent No. 6291654

GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.

TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: METTING, RAASCH & GEBHARDT, P.A.

STREET: 119 No. 6291654th Fourth Street, Suite 203

CITY: Minneapolis

STATE: Minnesota

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/308,022

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/20586

FILING DATE: 12-NOV-1997

APPLICATION NUMBER: 60/029,444

FILING DATE: 12-NOV-1996

APPLICATION NUMBER: 60/038,086

FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368  
FILING DATE: 19-SEP-1997  
APPLICATION NUMBER: 60/062,473  
FILING DATE: 16-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: MUEITING, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00430101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
308-022-6

Query Match 90.2%; Score 801.5; DB 3; Length 564;  
Best Local Similarity 86.2%; Pred. No. 2.5e-58;  
Matches 163; Conservative 13; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEARRADAK-----EKGKPKGRAGVPGELATPDKKENDAKSSDS 51  
DB 226 IKTDREAEAEARRADAKKEANVATSDGKPKGRAGVPGELATPDKKENDAKSSDS 285  
QY 52 SVGEETLPSPLKPEKKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIASDV 111  
DB 286 SVGEETLPSPLKSGKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIASDV 345  
QY 112 EVKKAELIYKEAEKPRNEEKVQAAYVESKKAETLEKIKTDKKAEEBAKRAAE 171  
DB 346 KVEAELEIYKEAEKPRDEKIKQAAYVESKKAETLEKIKTDKKAEEBAKRAAE 405  
QY 172 EDKVKKEPA 180  
DB 406 EDKVKKEPA 414

RESULT 8  
US-09-286-981B-18  
Sequence 18, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S.

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 18  
LENGTH: 406  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-18

Query Match 89.1%; Score 792; DB 4; Length 406;  
Best Local Similarity 90.2%; Pred. No. 1e-57;  
Matches 166; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY 1 IKTDREAEAEARRADAKQ---KPKGRAGVPGELATPDKKENDAKSSDSVGE 56  
DB 188 IKTDREAEAEARRADAKQDSSSKRKRGRVGRGDIQGEQATPDKKENDAKSSDSVGE 247  
QY 57 TLPSPLKPEKKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIASDV 116

DB 248 TLPSPLKPEKKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIASDV 307  
QY 117 ELEIYKEAEKPRNEEKVQAAYVESKKAETLEKIKTDKKAEEBAKRAAE 176  
DB 308 ELEIYKEAEKPRNEEKVQAAYVESKKAETLEKIKTDKKAEEBAKRAAE 367  
QY 177 EKPA 180  
DB 368 EKPA 371

RESULT 9  
US-09-286-981B-11  
Sequence 11, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S.

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 11  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-11

Query Match 89.1%; Score 792; DB 4; Length 425;  
Best Local Similarity 85.4%; Pred. No. 1.1e-57;  
Matches 164; Conservative 8; Mismatches 8; Indels 12; Gaps 1;

QY 1 IKTDREAEAEARRADAK-----EKGKPKGRAGVPGELATPDKKENDAKS 48  
DB 188 IKTDREAEAEARRADAKKEANVATSDGKPKGRAGVPGELATPDKKENDAKS 247  
QY 49 SDSSVGEETLPSPLKPEKKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIAE 108  
DB 248 SDSSVGEETLPSPLKPEKKVAEAEKKVAAEDOKEDRRNPPTNTYKTLLELEIAE 307  
QY 109 SDVEVKAELIYKEAEKPRNEEKVQAAYVESKKAETLEKIKTDKKAEEBAKRA 168  
DB 308 SDVKAELIYKEAEKPRNEEKVQAAYVESKKAETLEKIKTDKKAEEBAKRA 367  
QY 169 AAEEDVKKEPA 180  
DB 368 AAEEDVKKEPA 379

RESULT 10  
US-09-286-981B-10  
Sequence 10, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzmann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S.  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 10  
LENGTH: 414  
TYPE: PRT

us-09-298-523c-13\_copy\_263\_442.rai

;	CURRENT APPLICA
;	CURRENT FILING

```

: CURRENT APPLICATION NUMBER: US/09/286,991B
:
: CURRENT FILING DATE: 1999-04-06
:
: PRIOR APPLICATION NUMBER: US 60/085,743
:
: PRIOR FILING DATE: 1998-05-15
:
: NUMBER OF SEQ ID NOS: 38
:
: SOFTWARE: PatentIn Ver. 2.1

```

```

LENGTH: 412
TYPE: PRT
ORGANISM: Streptococcus pneumoniae

```

Query Match	87.2%	Score	775.5;	DB 4;	Length	412;
Best Local Similarity	89.6%	Pred.	No. 2.4e-56;			

[illegible]

58 LPSPSLKPEKKVAEAEKVEEAKKKAAEDOKEDRRNYPNTNYKTLLEIAESDVEVKKAE 117

118 LELVEAKEPRNEEKVQAKAEVSKKAATRLTKI KTDKSKAAEEAKRKAAEEDKVE 177  
311 LELVKEEAGSRNEPRKVKAKAEVSKKAATRLTKI KTDKSKAAEEAKRKAAEEDKVE 370

Qy	178	KPa	180
Qb	371	KPa	373

RESULT 13  
US-09-286-981B-14

Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Witzemann, Theresa M.

APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396

CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15

```

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 424

```

ORGANISM: *Streptococcus pneumoniae*  
US-09-286-981B-14

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

52 SVCEETI.PSPSI.PDEKKAFAEKVEEAKKA.FOKOEPEEPBNYPNTYVTL.PI.FTAESDV 111  
190 IKTDREKA-EEAKRRADAKLOEANVATSQDSKRRAKREVLGELATPPDKKENDDAKSSDS 248

249 SVGEETLTSPSLPEKRVAAEAEKKVVEEAKKKADQKEEDRNRYPNTYTLLELEIAESDV 308

309	EVAKAELELAKAEAKESKNEKKI KUVANVESKAEVUULEN KI DRKAAEEBANKKAA	368
171	EDDKVEKKA 180 	

;	CURRENT APPLICA
;	CURRENT FILING

14;	;	PRIOR FILING DATA
9;	;	NUMBER OF SEQ 1
Gaps	;	SOFTWARE: Patent
1;		

KKENDAKSSDS	51	LENGTH: 412
		TYPE: PRT
KKENDAKSSDS	248	ORGANISM: Stre

TLELEIAESDV	111	
:		
TLDELEIASDV	308	
		Query Match
		Best Local Simil

AABEAKRAAE	171
AABEAKRAAE	368
QY	1 IKTT
Dh	
	TYTT
	101

QY 58 LPSE  

b6  
b7C  
b7D

Qy	118	LELV
Db	311	LELV

Qy	178 KPA
Db	371 KPA

for Vaccines

---

RESULT 13  
US-09-286-981B-14

Patent No. 65035  
GENERAL INFORMAT  
APPLICANT: WIZ

APPLICANT: Job  
TITLE OF INVENT  
FILE REFERENCE:

```

; CURRENT FILING
;
; PRIOR APPLICATION
; PRIOR FILING DATA
;

```

```

6; Gaps 1;
; SOFTWARE: Pate
; SEQ ID NO 14
; LENGTH: 424
;

```

ORGANISM: Streptococcus  
US-09-286-981B-14

Best Local Simil  
Matches 166; C

AEEDKVEKPA 382	
Db	190 IKTD
Ov	52 SVGE

Db	249	SVGE
Ov	112	EVKK

for Vaccines	171	EEDK
	QY	
	DD	309 EVKK

Db 369 EEDKVEKPA 378

RESULT 14  
US-09-286-981B-12  
Sequence 12, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
ID NO 12  
LENGTH: 426  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-12

Query Match 86.7%; Score 770.5; DB 4; Length 426;  
Best Local Similarity 87.4%; Pred. No. 6.4e-56;  
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKRAKRGVPGELATPDKKENDAKSSDS 51  
Db 190 IKTDREKA-EAAGRRADAKLQEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDS 248  
QY 52 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 111  
Db 249 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 308  
QY 112 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDKKA-ESEAKRKA 170  
Db 309 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDKKA-ESEAKRKA 368  
QY 171 EEDKVEKPA 180  
Db 369 EEDKVEKPA 378

QY 15  
US-09-286-981B-13  
Sequence 13, Application US/09286981B  
Patent No. 6503511  
GENERAL INFORMATION:  
APPLICANT: Wismann, Theresa M.  
APPLICANT: Koenig, Scott  
APPLICANT: Johnson, Leslie S  
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines  
FILE REFERENCE: 469201-396  
CURRENT APPLICATION NUMBER: US/09/286,981B  
CURRENT FILING DATE: 1999-04-06  
PRIOR APPLICATION NUMBER: US 60/085,743  
PRIOR FILING DATE: 1998-05-15  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-286-981B-13

Query Match 86.6%; Score 769.5; DB 4; Length 425;  
Best Local Similarity 87.4%; Pred. No. 7.7e-56;  
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKRAKRGVPGELATPDKKENDAKSSDS 51

Db 190 IKTDREKA-EAAGRRADAKLQEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDS 248  
QY 52 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 111  
Db 249 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 308  
QY 112 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDKKA-ESEAKRKA 170  
Db 309 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDKKA-ESEAKRKA 368  
QY 171 EEDKVEKPA 180  
Db 369 EEDKVEKPA 378

Search completed: November 21, 2003, 13:35:11  
Job time: 8.55427 secs